Title

Mechanisms of RNA sorting into distinct extracellular vesicle sub-populations

Author

Temoche-Diaz, Morayma M

Publication Date

2020

Peer reviewed|Thesis/dissertation

Mechanisms of RNA sorting into distinct extracellular vesicle sub-populations

By

Morayma M Temoche-Diaz

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Randy Schekman, Chair
Professor Kathleen Collins
Professor Arash Komeili
Professor Donald Rio

Spring 2020

Abstract

Mechanisms of RNA sorting into distinct extracellular vesicle sub-populations

by

Morayma M Temoche-Diaz

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Randy Schekman, Chair

Extracellular vesicles (EVs) encompass a variety of vesicles secreted into the extracellular space. EVs have been implicated in promoting tumor metastasis, but the molecular composition of tumor-derived EV sub-types and the mechanisms by which molecules are sorted into EVs remain mostly unknown.

In the work described herein I used biochemical and genetic tools to fractionate distinct EV sub-populations and dissect their mechanisms of miRNA sorting. I report the separation of two small EV sub-populations from a metastatic breast cancer cell line, with biochemical features consistent with different sub-cellular origins. I then characterized their RNA content and observed that the EV sub-types use different mechanisms of miRNA sorting (selective and non-selective). Using a cell-free reaction I identified the Lupus La protein as the mediator of miR-122 *in vitro* packaging. I next showed that the La protein is required for the secretion of selectively sorted miRNAs and 5' TOP mRNAs *in vivo*. Finally by using proximity labeling I sought to understand how the La protein is recognized for sorting in the endosomal membrane. I found that the La protein is recognized for sorting by the sequestosome-1 protein. Its secretion is dependent on the presence of LC3-II in the endosomal compartment. In sum, this work provides a mechanistic understanding of miRNA sorting into EVs derived from cancer cells. Moreover I provide preliminary data on how RNA binding proteins might be recognized for secretion in endosomal compartments.

Table of Contents

About this dissertation

This dissertation describes the work performed in the laboratory of Dr. Randy Schekman, Professor in the Department of Molecular and Cellular Biology from 2015-2020. The main text is organized in four chapters. The first chapter is a revision on the current knowledge of extracellular vesicles, their heterogeneity, RNA cargo and fate. The next three chapters are organized in four sections (Introduction, Materials and Methods, Results and Discussions). Chapters 2, 3 and 4 describe primary research findings in relation to extracellular vesicle sub-populations and their mechanisms of RNA packaging. Chapter 2 describes the purification of distinct EV sub-populations and characterization of their RNA content. Chapter 3 describes the findings in relation to the RNA binding proteins mediating selective miRNA sorting. Chapter 4 seeks to understand how the RNA binding proteins are recognized for secretion in the endosomal compartments.

Acknowledgements

I would like first to thank Randy Schekman for being my mentor for these past 5 years. Having you as role model for the kind of scientist I wanted to become has changed me academically and I am infinitely grateful for that.

Thank you to Kathleen Collins, Arash Komeili and Donald Rio. I am extremely grateful for your invaluable academic advice and all the personal support.

Thank you to Alan Lambowitz for collaborating with me and helping generating the sequencing data on Chapter 2 and 3.

Thank you to the staff at UC Berkeley Tissue Culture Facility, the DNA sequencing facility, the Vincent J Coates Proteomics Facility and the Flow Cytometry Facility.

Thank you Matthew Shurtleff for being a great mentor, benchmate and friend. Also thank you to all my previous academic mentors: Justin McDonough, Craig Roy, Michael Talledo and Jorge Arevalo. Thank you for paving the way of my academic life.

Thank you to my lifelong friends Diego, Amanda and Matt. Thank you for always being with me on my ups and downs.

Thank you Thibault for your love and support.

Finally and most importantly thank you to my family: my parents and siblings. Specially thank you to the two women who raised me and made me who I am: Morayma Dias Chacon and Silvia Lopez Tejada. This is for you.

Chapter 1: Extracellular vesicles: heterogeneity, RNA cargo and their fate

Introduction

During growth, mammalian cells secrete structures termed extracellular vesicles (EVs) into the extracellular environment that can be isolated from the conditioned medium of cultured cells and have been detected in all bodily fluids (*1-4*). EVs are membranous compartments that consist of a lipid bilayer with a unique set of transmembrane proteins enclosing soluble contents that include nucleic acids and proteins (*5*). Cells release an array of EV sub-populations with different sub-cellular origins and distinctive protein (*6*) and RNA (*7, 8*) cargo. Their physiological role remains a debate in the field.

In this section I first describe what is known about EV heterogeneity, then I focus on revising the current knowledge of EV nucleic acid cargo and finally I discuss the possible routes of EV fate.

EV heterogeneity and biogenesis

Cells release an array of EV sub-populations, which can be broadly classified into two categories: exosomes and shedding vesicles, distinguished by the cell membrane of origin. Exosomes are 30-150 nm vesicles that originate in the endocytic pathway. Their secretion to the extracellular space occurs upon multivesicular body (MVB) fusion with the plasma membrane resulting in the release of intraluminal vesicles (ILVs) to the extracellular space (*9, 10*). Shedding vesicles, or microvesicles, refer to a more heterogeneous group of EVs, with sizes ranging from 30 to 1,000nm, which originate by budding directly from the plasma membrane (*11*).

Exosomes originate in the endocytic pathway. The endocytic pathway is a dynamic membrane trafficking process that consists of internalized macromolecules and membrane proteins which then recycle back to the plasma membrane/extracellular space or degrade them into the lysosome (*12*). During the process of maturation, from early to late endosomes, ILVs are created and accumulate into late endosomes – also called MVBs. Once MVBs fuse with the plasma membrane the secreted ILVs are now termed exosomes (*5*). Thus understanding ILV biogenesis is crucial to understand exosome biogenesis. To date two main pathways of ILV/exosome biogenesis have been described. The first pathway requires the endosomal sorting complex required for transport (ESCRT). There are four core subunits of the ESCRT complexes termed ESCRT-0, ESCRT-I, ESCRT-II and ESCRT-III (*13*). The first complex, ESCRT-0, is responsible for recognizing and clustering ubiquitinated cargo in the MVB. ESCRT-I and ESCRT-II subsequently are responsible for budding the limiting membrane of MVB into its lumen. ESCRT-III is required for the last step in ILV biogenesis, vesicle abscission. A second and less understood pathway of ILV/exosome biogenesis is ESCRT independent. This pathway requires the lipid ceramide to induce vesicle curvature and budding (*14*). Trajkovic et al. (2008) showed that ceramide biogenesis is required for ILV/exosome

biogenesis. Knocking down or inhibiting neutral sphingomyelinase 2 (nSMase2), enzyme responsible for hydrolysis of sphingomyelin into ceramide, was associated with an impaired exosomal release. A more recent study also links ceramide to ILV/exosome biogenesis (*15*). Leidal et al. (2020) demonstrated that lipidated LC3 (LC3-II) in the MVB membrane recruits the neutral sphingomyelinase SMPD3. SMPD3 is in turn responsible for ceramide biogenesis, thus promoting ILV/exosome biogenesis. How the cargo is recognized and recruited to the site of biogenesis in the ceramide-dependent pathway is unknown. I show data and discuss a possible route for this process in *Chapter 4*. Whether both ESCRT and ceramide dependent ILV/exosome biogenesis pathways act simultaneously in the same cells or are cell-specific is unknown. However it is plausible to assume that both pathways act simultaneously in the same cells. Abrogation of either one does not cause a complete abolishment on EVs secretion, but rather a partial reduction on their secretion (*14-16*).

ESCRT complexes and ceramide are required for ILV biogenesis. Exosomal secretion occurs when ILV-filled MVBs fuse with the plasma membrane. This fusion process most likely requires intracellular transport machinery, creation of docking sites between MVBs and plasma membrane and SNARE proteins to promote MVB-plasma membrane fusion. Three proteins involved in MVB-plasma membrane docking have been reported so far: Rab11a (*17*), Rab27a (*18*) and Rab35 (*19*). Fusogenic SNARE proteins involved in exosomal release have not yet been elucidated.

Shedding vesicles have their origin in the plasma membrane. The molecular mechanism behind their biogenesis is less well characterized. One study showed an interplay between the small GTPase ARF6 and the actin-myosin machinery to promote vesicle budding and abscission (*20*). Moreover the requirement of the ESCRT complex to mediate microvesicle biogenesis has also been reported. The adaptor protein arrestin domain-containing protein 1 (ARRDC1) has been shown to promote the re-localization of the ESCRT-I protein TSG101 from endosomes to plasma membrane and thus mediate ESCRT-dependent microvesicle biogenesis (*21*). Thus some of the machinery involved in microvesicle and ILV/exosome biogenesis might be shared.

Exosomes tend to be smaller (30-150 nm) and more homogenous than shedding vesicles (30-1,000 nm). The size of exosomes is restricted by the size of MVBs; such a restriction does not apply for shedding vesicles. As such, shedding vesicles and exosomes are also referred to as large and small EVs, respectively. Their isolation is generally achieved by differential ultracentrifugation. Evidence, however, that multiple EV species are being co-isolated as small EVs has arisen, proving that the small EV fraction represents a crude mix of different EV species. Studies further purifying different small EVs sub-populations have shown that they carry different protein (*6*) and RNA (*8*) signatures. Moreover, distinct small EV sub-populations have been implicated in mediating various physiological responses (*22*). In *Chapter 2* I describe a methodology to purify distinct small EV sub-populations. I show data confirming that the material isolated as small EV contains not only bona fide exosomes, but also small microvesicles likely budding from the plasma membrane. Studies purifying distinct small EV sub-populations are needed in order to understand their mechanisms of biogenesis/cargo

sorting and ultimately their physiological role.

EV associated nucleic acid

DNA

The presence of DNA in the high-speed crude pellet (*23-25*), in vesicles isolated by buoyant density separation on sucrose (*26*) and Optiprep gradients (own unpublished observations) has been reported. Shurtleff et al. (2017) showed that the presence of DNA associated with EVs is largely due to DNA associating with the surface of vesicles (*26*). Libraries for RNA sequencing were obtained from EVs floated on a sucrose step gradient. Buoyant EVs were treated with DNase in the presence or absence of non-ionic detergent prior to nucleic acid extraction and library preparation. Reads mapping to protein coding genes, specifically to intronic regions, were greatly reduced upon the addition of DNase alone. A similar effect was observed for reads mapping to long intergenic noncoding RNAs, antisense RNAs and other lncRNAs. Thus DNA is not a bona fide EV cargo, its presence on EV preparations is due to contaminant DNA associated with the EV surface. Other studies suggest that as much as 50% of DNA associated with EVs resides inside the vesicles as shown by DNase protection assays (*24, 25*). However these studies lack proper controls to confidently affirm that the non-degraded DNA is inside vesicles. EV membrane permeabilization with non-ionic detergent and/or evidence that the nucleases are capable of fully degrading the same amount of DNA with the conditions used is an essential control.

A recent study confirmed the presence of DNA in high-speed crude pellets. The DNA signal was greatly reduced when EVs were purified by flotation on Optiprep linear gradients (*23*). The mechanism of DNA secretion was then studied. Jeppesen et al. (2019) showed that micronuclei co-localized with CD63 positive MVBs, but their peak intensity was offset, indicating that DNA did not co-localize with CD63 positive ILVs. Jeppesen et al. (2019) propose an amphisome-mediated mechanism for DNA secretion. Cytoplasmic micronuclei might be engulfed by an autophagosome. DNA-containing autophagosomes in turn may fuse with CD63 positive MVBs to form amphisomes. The amphisomes then may fuse with the plasma membrane to secrete cytoplasmic DNA. Consistent with this proposed mechanism of DNA secretion, DNA is present in the high-speed crude pellet, but as vesicle-free species.

RNA

The first report of the presence of RNA associated with EVs occurred over a decade ago (*27*). Valadi et al. (2007) demonstrated an enrichment of small RNA species in the high-speed crude pellet derived from conditioned medium from human mast cell culture in comparison to cell lysates. Importantly the presence and enrichment of certain miRNAs in the high-speed pellet was also noted. Since then miRNAs have garnered special attention in the EV field. MiRNAs are ~22 nucleotide transcripts that modulate gene expression at the post-transcriptional level (*28*). Although the functional importance of EV-miRNAs, as signaling molecules has received some support, including in immunologic response and metastatic tumor cell growth (*29-39*), the molecular mechanisms and regulation of sorting miRNAs into EVs are not well understood.

Non-selective miRNA sorting has been reported to occur in EVs that originate in the plasma membrane (*7, 8*). Non-selectivity here is defined as the lack of EV enrichment in comparison to intracellular lysates. Whereas selectivity in miRNA sorting has been shown to occur in in bona fide exosomes, as shown by CD63 vesicle immunoprecipitation (IP) (*40*) and fractionation of EV sub-populations in a high-resolution Optiprep linear gradient (*8*). Importantly Ago2 has not been detected in pure EV samples (*8, 23, 40, 41*), thus other RNA binding proteins (RBPs) must be associated with miRNAs and mediate their sorting into EVs. To date four RBPs have been shown to mediate the selective mechanism of miRNA sorting into EVs derived from human cells: hnRNPA2B1 (*42*), SYNCRIP (*43*), YBX1(*40*) and the La protein (*8*). It is thought that the binding of these RBPs is at least partially dependent on short motifs present in their cognate miRNAs (*8, 42, 43*). Importantly these four studies showed that the RBP-miRNA complexes are secreted into the extracellular space. This suggests that the RBPs mediating miRNA sorting are not recycled back to the cytoplasm. The miRNA-RBP complex might represent a high-affinity interaction. Villaroya-Beltri et al. (2013) showed that sumoylation is required for hnRNPA2B1 secretion into EVs. Whether sumoylation or other post-translational modification is required for the secretion of SYNCRIP, YBX1 and the La protein is unknown.

Despite the interest of the EV science community in miRNAs, miRNAs represent only a small portion of the total pool of EV transcripts. When a thermostable group II intron reverse transcriptase (TGIRT) was used to create comprehensive EV cDNA libraries, miRNAs represented less than 4% of the total pool of small non-coding RNAs (*8, 26*). Data obtained by TGIRT-seq showed that the most common EV small non-coding transcript is full-length tRNA, representing at least 80% of the total small non-coding RNA pool. The reported full-length tRNA representation does not seem to be a overestimation as a result of potential biases of TGIRT-Seq. Multiple studies have reported Bioanalyzer data for EV RNAs confirming the TGIRT-seq observations. These studies report a predominant peak corresponding to full-length tRNA as shown by Bioanalyzer fractionation (*8, 23, 26, 27, 31, 39, 44-46*). Many studies performing RNA sequencing in EV samples have reported the presence of predominantly tRNA fragments as EV residents (*7, 45-48*). These reports might be an artefact in the use of conventional reverse transcriptases which bias against highly modified or structured RNAs, such as tRNAs. Moreover the under-representation of full length mature tRNAs can also be the result of biases in the bioinformatics analysis. Transcripts with moderate mismatches to the reference genome are generally discarded throughout the analysis. The loss of full length tRNA reads derived from EVs by bioinformatics analysis due to low lenience of mismatch has been observed (Kathleen Collins, personal communication). Importantly a recent study has demonstrated the presence of full-length tRNA in EVs purified by Optiprep flotation by using DIG-based Northern blot (*49*). Tosar et al. (2020) demonstrated that tRNA lysine UUU is predominantly present as a full-length species in purified EVs. In *Chapter 2,* I demonstrate that tRNA lysine UUU is a highly EV enriched transcript and it is amongst the most abundant in isolated EVs. Moreover Tosar et al. (2020) demonstrated that when cells are grown in the presence of exo-depleted FBS, most of the tRNA associated with EVs is full-length. A high enrichment of tRNA fragments sedimentable at high-speed spins is observed when cells are grown under

serum-free medium. This could be as a result of the absence of FBS-nucleases available to degrade the EV-free tRNA and/or due to physiological changes under starvation. Thus, the reports of tRNA fragments in EVs might partially be due to the use of conventional reverse transcriptases, low lenience of mismatch in bioinformatics analysis and the use of serum-free conditions to harvest conditioned medium.

The presence of YRNA (*8, 15, 23, 26, 46, 50*) and 7SL (*8, 26, 46, 51*) in EVs has been demonstrated by RNA-sequencing. Reports about their full-length nature are supported by northern blot analysis (*49, 50, 52*). Interestingly the presence of both, YRNA and 7SL has been reported in virions of RNA viruses (*53, 54*). There might be a conserved mechanism by which RNA viruses and EVs package these cellular RNAs.

Besides the small non-coding RNAs mentioned above, the presence of mRNA in EVs has also been reported (*26, 27, 55, 56*). Interestingly an enrichment of 5' terminal oligopyrimide mRNAs (5' TOP mRNAs) has been observed in EVs in comparison to cellular lysates (*26*). 5' TOP mRNAs encode ribosomal proteins and translation elongation factors. They are translationally repressed when cells encounter hostile growth conditions such as starvation (*57*). In *Chapter 3,* I show data consistent with a requirement of the La protein for 5' TOP mRNA secretion.

Some of the RBP players involved in the sorting of miRNAs and other transcripts into EVs are known. However the bigger picture of how RBPS and their associated transcripts are sorted into EVs is still not well understood. Temoche-Diaz et al. (2019) showed that the La protein, responsible for selective miRNA secretion into EVs, forms cytoplasmic puncta that associates with late endosomes. Thus RNA granules might be involved in EV-mediated RNA secretion. A recent study showed that RNA granules hitchhike on lysosomes for transport using annexin A11 as their tether (*58*). This opens up the possibility that RNA granules already tethered on lysosomes or late endosomes are readily available for their recognition and sorting into ILVs. In fact, knocking down annexin A2 has been shown to reduce EV-associated miRNA secretion and its overexpression to promote miRNA secretion (*59*). It is possible that annexin A2 might have a similar function as annexin A11 by tethering RNA granules into late endosomes. More research in this area will prove helpful to understand the mechanisms of RBP-RNA loading into EVs.

Extracellular Vesicle Fate

The fate of EVs upon release is uncertain. Its remains possible that most EVs serve the purpose of disposing of unwanted cellular constituents by secretion to the extracellular space, where they would remain until they are damaged and degraded or internalized into the lysosome of a phagocytic cell. Alternatively, EVs could fuse with recipient cells at the plasma membrane or after internalization at an endosomal membrane and, in so doing, serve a regulatory function. Such functional transfer of EV proteins (*60-62*), mRNAs (*27, 55, 56*) and miRNAs (*30-32, 34-39, 63*) has been suggested. For transfer of soluble constituents, some form of membrane fusion will be essential; however, there is scarce evidence of a role for a surface receptor or a fusion protein either in the EV or the recipient cell, with few exceptions such as Christianson & Belting's

(2014) evidence that heparin sulfate proteoglycans are important cell receptors for EV internalization (*64*).

Despite the broad interest of the scientific community, studies giving direct evidence of EV-mediated information transfer are rare. One notable study provided evidence of direct transfer of EV content to a recipient cell using a Cre-loxP reporter system (*65*). With a reporter cell line that switches fluorescence from red to green upon delivery of Cre, Zomer et al. (2016) visualized a low level of information transfer during two-week co-culture experiments. The effect was attributed to transfer of Cre mRNA rather than protein because the mRNA could be detected by reverse transcription polymerase chain reaction (RT-PCR) in crudely isolated EVs, but Cre protein was undetectable by immunoblot. Unfortunately, the claim was not accompanied by a quantitative evaluation of the limit of detection for immunoblot versus RT-PCR. Furthermore, no evidence of Cre transfer by purified EVs was reported. More convincingly, Lai et al. (2015) showed that Gaussia luciferase mRNA could be transferred through EVs and translated in recipient cells (*56*). Expression of the transferred mRNA was sensitive to cycloheximide with a significantly lower luminescence signal compared to nontreated control. Follow-up studies with purified EVs carrying luciferase mRNA would be instructive.

Most of the functional studies to date suggest that cells internalize EVs through endocytosis (*30, 64, 66-70*). When claiming internalization or uptake of EVs into recipient cells, investigators must be aware that internalization may simply lead to destruction of EV content in the lysosome without release of luminal content into the cytoplasm. Functional delivery of soluble content requires crossing the membrane barrier by means of EV-cell membrane fusion. Functional delivery may follow a path taken by enveloped viruses that fuse at the cell surface or in the acidic environment of the endosome (*71, 72*). It is noteworthy that intraluminal vesicles have been shown to back-fuse with the luminal endosomal membrane surface of a multivesicular body (*73, 74*), suggesting that EV-endosomal compartment fusion is a feasible means of EV content delivery to the cytoplasm. Despite the literature suggesting EV uptake, fusion at the plasma membrane may be a prominent means of content delivery to the cytoplasm. EVs are enriched for tetraspanin family proteins including CD9, a transmembrane protein required for egg-sperm fusion (*75, 76*), suggesting that tetraspanins might play a role in EV-plasma membrane fusion. Determining unambiguously the site of EV entry will be crucial to understanding the mechanisms of EV-mediated intercellular communication.

Studies showing internalization of EVs into recipient cells have employed exogenous lipophilic dyes to stain partially purified preparations of EVs (*62, 77-79*) or transmembrane proteins tagged with fluorescent reporters (*67, 70, 80*). Internalization has been evaluated by flow cytometry or fluorescence microscopy. When using lipophilic dies, there are two major points to consider. First, dyes might change the nature of EVs, increasing or decreasing their internalization. Secondly, dyes may produce false positives due to their intrinsic chemical properties, e.g., a longer half-life and a tendency to aggregate (*56, 81*). Furthermore, as discussed above, the purity of an EV preparation bears directly on any interpretation based on a bulk labeling approach. For these reasons, it is preferable to tag EV membranes, either by fusing EV transmembrane proteins to

fluorescent reporters (*67, 70*), or by directing fluorescent reporters to membranes by the use of palmitoylation signals (*56, 82*). Experiments built on following EV fate based on membrane fluorescence may address whether EVs are internalized into cells; however, an evaluation of content delivery requires further analysis. Discerning if membrane fusion and therefore content delivery have happened requires tagging both the membrane and soluble EV content. EV fusion with a small intracellular organelle, such an endosome, will lead to loss of soluble internal EV content signal (due to dilution by diffusion into the cytoplasm), without the disappearance of the membrane marker (due to restricted diffusion into the only slightly larger endocytic membrane), as shown in experiments to determine the location of HIV entry (*83*). This approach applied to EVs could document fusion with the endosome and permits the process to be quantified. Although such fusion has been suggested, many groups have reported that EVs colocalize with lysosomal markers in recipient cells (*67, 69, 80*), likely resulting in EV content degradation. Elegant experiments done by Heusermann et al. (2016) have shown that up to 60% of internalized CD63-positive EVs co-localize with lysosomes after two days of exposure, and no evidence of membrane fusion within endocytic vesicles was observed by transmission electron microscopy. However, different EV subpopulations and specific, targeted delivery of EVs to unique destinations to produce a functional response must be identified for closer inspection of the fate of internalized cargo content.

The major function of EVs may be, as originally proposed for their role in erythroid precursor cell maturation (*84*), the disposal of unnecessary or toxic membrane and soluble content. Secretion of cytoplasmic content via EVs might depend on an overabundance of specific molecules in producing cells. Squadrito et al. (2014) showed that the pool of EV-secreted miRNAs depends on their abundance and the abundance of their target mRNA in the producing cells. When miRNAs were in excess of their target mRNA, they were released in EVs, suggesting that cells may use EVs to purge excess material (*80*). Moreover, aggressive cancer cells secrete tumor-suppressor miRNAs in EVs possibly as a means to maintain the metastatic state (*85*). Of note is the apparent absence of Ago2 in EVs from some cell types (*8, 23, 40, 41, 85*). Free miRNAs interact with membrane proteins and, perhaps as a result, impact cellular processes (*86*). This raises the possibility that cells sort RNA-induced silencing complex (RISC)-dissociated miRNAs into EVs to remove them from the cytoplasm, thus alleviating potential deleterious effects and freeing up RISC machinery for miRNAs with abundant targets. Moreover the presence of YRNAs in EVs might suggest that EV-mediated RNA secretion acts as mechanism of RNA quality control. YRNA forms a complex with the protein Ro60. The Ro60-YRNA complex promotes RNA quality control in bacteria by associating with an exonuclease, thus forming a RNA degradation machine (*87*). Furthermore in eukaryotes the Ro60-YRNA complex has been reported to bind misfolded RNAs (*88*) and presumably act in RNA quality control (*89*). Thus the YRNA-Ro60 complex might recognize misfolded, damaged, non-functional or in excess transcripts and target them for disposal via an EV-mediated secretion.

Importantly for this thesis, in the context of cancer biology, cancer cells are known to produce more EVs with distinct content than their nonmalignant counterparts (*90*). This abundance may be a result of EV's evolved function in the spread of malignancy or

a cellular response to dispose of toxic material. The existence of different EV subpopulations produced by the same cell line is now largely accepted (*6, 8, 22, 91*), raising the possibility that different EV subpopulations might have different purposes or fates. Studies focusing on the mechanisms of sorting and the fate of discrete EV subpopulations will help to elucidate their biological functions.

Chapter 2: Fractionation of different EV sub-populations and characterization of their RNA content

Introduction

The existence of different subtypes of extracellular vesicles with different sub-cellular origins has been documented (*5*). Microvesicles are vesicles derived from the plasma membrane (*11*); their sizes can range from 30-1,000 nm. Exosomes are extracellular vesicles of endosomal origin and range from 30-150nm (*9, 10, 84*). Some microvesicles and exosomes sediment at medium-speed and high-speed during differential ultracentrifugation. The high-speed pellet has been long used as *exosomal* material for physiological and characterization studies, however the presence of other particulate material of non-exosomal origin in the high-speed pellet has been speculated (*6*). Therefore an in depth study of the particulate material sedimenting at high-speed is needed in order to better understand the mechanisms of cargo sorting and ultimately their physiological roles.

In order to fractionate the particulate material sedimenting at high-speed, I developed a two-step methodology consisting of differential ultracentrifugation followed by buoyant density flotation. I used the highly metastatic breast cancer cell line MDA-MB-231 as model cell line as numerous studies linked EVs to cancer metastasis. I identified that the high-speed pellet fraction consists of at least three different particulate species: RNA-protein complexes not associated with a lipidic membrane and two different extracellular vesicle sub-populations that equilibrate to slightly different buoyant densities. By using genetic and cell biology tools I found that the distinct EV sub-populations show characteristics suggesting different sub-cellular origins. By performing miRNA profiling, RNA sequencing and qRT-PCR, I showed that the distinct EV sub-populations have differential miRNA cargo, seemingly associated with their membrane of origin.

Materials and Methods

Cell lines, media and general chemicals

MDA-MB-231 cells were cultured in DMEM with 10% FBS (Thermo Fisher Scientific, Waltham, MA). MDA-MB-231 cells were confirmed by short tandem repeat profiling (STR) and tested negative for mycoplasma contamination. For EV production, cells were seeded at ~10% confluency in 10% exosome-depleted FBS (System Biosciences, Palo Alto, CA) in 150mm CellBIND tissue culture dishes (Corning, Corning, NY) containing 30ml of growth medium. EVs were collected when cells reached ~80% confluency (~72h). Unless otherwise noted, all chemicals were purchased from Sigma Aldrich (St. Louis, MO).

Extracellular vesicle purification

Conditioned medium (720 ml for TGIRT-seq and miRNA profiling and 420 ml for all other experiments) was harvested from 80% confluent MDA-MB-231 cultured cells. All subsequent manipulations were performed at 4°C. Cells and large debris were removed by centrifugation in a Sorvall R6+ centrifuge (Thermo Fisher Scientific) at

1,000xg for 15 min followed by 10,000xg for 15 min in 500 ml vessels using a fixed angle FIBERlite F14-6X500y rotor (Thermo Fisher Scientific). The supernatant fraction was then centrifuged at ~100,000xg (28,000 RPM) onto a 60% sucrose cushion in buffer C (10mM HEPES pH 7,4, 0.85% w/v NaCl) for 1.5 h using two SW-28 rotors. The interface over the sucrose cushion was collected and pooled for an additional ~130,000xg (32,500 RPM) centrifugation over a 60% sucrose cushion in a SW41 rotor for 15 h. The collected interface from the first sucrose cushion should not exceed a sucrose concentration of 21%, as measured by refractometry, for the second centrifugation in the SW41 to be successful. Higher concentrations of sucrose cause the EVs to equilibrate at the ambient buoyant density, impeding sedimentation. For purification of EV sub-populations based on their distinct buoyant density, the cushion-sedimented vesicles were collected and mixed with 60% sucrose to a final volume of 4ml (sucrose final concentration ~40%). Layers of 1.5 ml of 25%, 20%, 15%, 10% and 5% iodixanol (Optiprep) solutions in buffer C were sequentially overlaid and samples were centrifuged at ~160,000xg (36,000 RPM) for 15 h in a SW41 rotor. Fractions (400ul) from top to bottom were then collected and mixed with Laemmli buffer for immunoblot analysis or RNA was extracted using a mirVana miRNA isolation kit (Thermo Fischer Scientific). In some cases, such as the detection of CD63, La or nucleolin, the floated fractions were centrifuged in a SW41 rotor at ~130,000g in order to concentrate them and improve their detection. This method produced a linear density gradient from 1.036 to 1.24 g/ml.

Mass spectrometry analysis of EV sub-populations

EV sub-populations purified from the iodixanol gradient (as described above) were diluted with buffer C to a final volume of 30ml. The diluted samples were then centrifuged at ~100,000xg (28,000RPM) in a SW28 rotor for 1.5 h. Pellet samples were resuspended in Laemmli buffer without bromophenol blue and electrophoresed in a 4-20% acrylamide Tris-Glycine gradient gel (Life Technologies) for ~3 min. The bulk of proteins were stained with Coomassie and the stained band was excised from the gel with a new razor blade. Samples were submitted to the Vincent J. Coates Proteomics/Mass Spectroscopy laboratory at UC Berkeley for in-gel tryptic digestion of proteins followed by liquid chromatography and mass spectrometry analysis according to their standards.

For sub-cellular localization analysis of the detected proteins, the list of proteins detected by mass spectrometry was analyzed first using the GoTermFinder developed at the Lewis-Sigler Institute at Princeton (*92*) followed by REVIGO analysis (*93*) to obtain the scatterplots of sub-cellular localization along with their organelle association.

Nanoparticle tracking analysis

Extracellular vesicles purified on linear iodixanol gradients were diluted 1:100 with PBS filtered with a 0.02um filter (Whatman GmbH, Dassel, Germany). The mixture was drawn into a 1 ml syringe and inserted into a Nanosight LM10 instrument (Malvern, UK). Particles were tracked for 60 s using Nanosight nanoparticle tracking analysis software. Each sample was analyzed 5 times and the counts were averaged.

Immunoblots

Exosomes were prepared by mixing sedimented vesicles with 1X Laemmli buffer,

or 1X Laemmli buffer without DTT when detection of CD63 was performed. Cell lysates were prepared by adding lysis buffer (10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 0.1% sodium dodecyl sulfate, 0.5% sodium deoxycholate, 1% Triton X-100, 10% glycerol) to cell pellets. Protein was quantified using a BCA Protein Assay Kit (Thermo Fischer Scientific), and the selected amount was mixed with Laemmli buffer. Samples were heated at 95°C for 5 min and separated on 4-20% acrylamide Tris-Glycine gradient gels (Life Technologies). Proteins were transferred to polyvinylidene difluoride membranes (EMD Millipore, Darmstadt, Germany), blocked with 5% bovine serum albumin in TBST and incubated overnight with primary antibodies. Blots were then washed with TBST, incubated with anti-rabbit or anti-mouse secondary antibodies (GE Healthcare Life Sciences, Pittsbugh, PA) and detected with ECL-2 reagent (Thermo Fisher Scientific). Primary antibodies used in this study were anti-CD9 #13174S (Cell Signaling Technology, Danvers, MA), anti-alpha integrin alpha 3 #ab190731 (Abcam, Cambridge, MA), anti-Mfg-e8 #MAB2767 (R&D systems, Minneapolis, MN), anti-flotillin-2 #610383 (BD Biosciences, San Jose, CA), anti-CD63 #BDB556019 (BD Biosciences), anti-Ago2 #2897 (Cell Signaling Technology), anti Ago2 ab#186733 (abcam), anti-Dicer #sc-30226 (Santa Cruz Biotechnology), anti-vinculin #ab129002 (Abcam), anti-Rab27a ab55667 (Abcam), anti-Rab35 #9690S (Cell Signaling)!

Quantitative real-time PCR

Cellular and EV RNAs were extracted with a mirVana miRNA isolation kit (Thermo Fischer Scientific), unless otherwise specified. Taqman miRNA assays for miRNA detection were purchased from Life Technologies. Assay numbers were: hsa-miR-193b, #002366; hsa-miR-29b-1, #002165; hsa-miR-486, #001278; hsa-miR-574-3p, #002349; hsa-miR-320a, #002277; hsa-miR-142-3p, #000464; hsa-miR-126, #000451; hsa-mir-145-5p, #002278; hsa-mir-122-5p, #002245; hsa-mir-451a, #001141; hsa-mir-182-5, 002334; hsa-miR-429, #001024; hsa-miR-675, #121124_mat; hsa-miR-193a-5p, #002281, hsa-miR-185 #002271. As there are no well-accepted endogenous control transcripts for EVs, the relative quantification was normalized to equal amounts of starting RNA material. RNA was quantified by an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) according to the manufacturer's instructions. RNA (1ng) was used for reverse transcription according to the manufacturer's instructions. Relative quantification was calculated from the expression $2^{-(Ct_{(control)}-Ct_{(experimental)})}$. Taqman qPCR master mix with no amperase UNG was obtained from Life Technologies and quantitative real-time PCR was performed using an ABI-7900 real-time PCR system (Life Technologies).

Immunofluorescence

MDA-MB-231 cells on 12mm round coverslips (Corning) were fixed by adding 4% EM-grade formaldehyde (Electron Microscopy Sciences, Hatfield, PA) for 20 min at room temperature. Subsequently, cells were washed 3 times with PBS and permeabilized/blocked by adding blocking buffer (0.1% TX-100 in 2 FBS% for 20 min. Cells were then incubated with 1:100 dilution of CD9 (Biolegend, #312102) and 1:100 dilution of CD63 (DSHB, #H5C6) in blocking buffer for 1.5 h at room temperature, extensively washed and incubated in secondary antibodies diluted 1:1,000 in blocking buffer and Alexa Fluor 488 (Thermo Fischer Scientific), for 1.5 h. Cells were extensively

washed, rinsed briefly in dH_20 and mounted on slides with ProLong Gold with DAPI (Thermo Fischer). Cells were imaged keeping all the settings constant in an Axio Observer Z1 (Zeiss, Oberkochen, Germany). For superresolution microscopy, images were taken with an Elyra P.1 (Zeiss)

miRNA profiling

I used the Discovery Panel from Firefly Service at Abcam to profile 408 miRNAs. Aliquots of 40 µl of 270pg/µl of RNA (samples included HD, LD and CL) were shipped to the Abcam service and the profiling was done according to their standards. Samples were normalized to total amount of RNA. The data used for the analysis was raw data with background subtracted. Only data with more than 0.1 arbitrary units of fluorescence (AUF) was considered for further analysis.

RNA extraction and TGIRT-seq library preparation

Cellular and EV RNAs were extracted by using a mirVana miRNA Isolation Kit. Cellular RNA was extracted using a modified version of the mirVana protocol, enriching for <200nt transcripts, in order to reflect EV RNA composition by transcript size. EV RNA was extracted using the standard miRVana protocol. TGIRT-seq libraries were prepared from 2-10ng of starting material.

TGIRT-seq libraries were prepared essentially as described (*94*), with a modification in the starting molecule (see below). Reverse transcription reactions contained purified RNAs, buffer (20 mM Tris-HCl, pH7.5, 450 mM NaCl, 5 mM $MgCl_2$), 5 mM DTT, 100 nM starting molecule (see below) and 1 µM TGIRT-III (Ingex). Reverse transcription by TGIRT-III is initiated by template switching from a starting molecule consisting of a DNA primer (5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATTAN-3') encoding the reverse complement of the Illumina Read2 sequencing primer binding site (R2R) annealed to a complementary RNA oligonucleotide (R2) such that there is a single nucleotide 3' DNA overhang composed of an equimolar mixture of A, G, C and T. The RNA oligonucleotide is blocked at its 3' end with C3Sp (IDT) to inhibit template switching to itself. Reactions were pre-incubated at room temperature for 30 min and then initiated by addition of 1 mM dNTPs. Reactions were then incubated at 60°C for 15 min and terminated by adding 5 N NaOH to a final concentration of 0.25 N and incubated at 95°C for 3 min to degrade RNAs and denature protein. The reactions were then cooled to room temperature and neutralized with 5 N HCl. cDNAs were purified by using a Qiagen MinElute Reaction Cleanup Kit and then ligated at their 3' ends to a modified DNA oligonucleotide (5'Phos-CCTGTTATCCCTAGATCGTCGGACTGTAGAACTCTGAACGTGTAC-3'-C3Sp) encoding the reverse complement of the Illumina Read1 primer binding site (R1R) using Thermostable 5' AppDNA/RNA Ligase (New England Biolabs). Ligated cDNAs were re-purified with MinElute Reaction Cleanup Kit and amplified by PCR for 12 cycles using Phusion DNA polymerase (Thermo Fisher Scientific) with overlapping multiplex and barcode primers that add sequences necessary for Illumina sequencing. PCR reactions were purified using a Select-a-size DNA Clean and Concentrator (Zymo) to remove adapter dimers. Libraries were sequenced on a NextSeq 500 instrument (75-nt,

single end reads) at the Genome Sequencing and Analysis Facility at the University of Texas at Austin.

RNA sequencing analysis

Illumina TruSeq adapters and PCR primer sequences were trimmed from the reads with cutadapt 1.16 (*95*) (sequencing quality score cut-off at 20) and reads <15-nt after trimmings were discarded. Reads were then mapped using HISAT2 v2.0.2 (*96*) with default settings to the human genome reference sequence (Ensembl GRCh38 Release 76) combined with additional contigs for 5S and 45S rRNA genes and the *E. coli* genome sequence (Genebank: NC_000913) (denoted Pass 1). The additional contigs for the 5S and 45S rRNA genes included the 2.2-kb 5S rRNA repeats from the 5S rRNA cluster on chromosome 1 (1q42, GeneBank: X12811) and the 43-kb 45S rRNA repeats that contained 5.8S, 18S and 28S rRNAs from clusters on chromosomes 13,14,15,21, and 22 (GeneBank: U13369). Unmapped reads from Pass 1 were re-mapped to Ensembl GRCh38 Release 76 by Bowtie 2 v2.2.6 (*97*) with local alignment to improve the mapping rate for reads containing post-transcriptionally added 5' or 3' nucleotides (e.g., CCA and poly(U)), short untrimmed adapter sequences, or non-templated nucleotides added to the 3' end of the cDNAs by the TGIRT enzyme (denoted Pass 2). The mapped reads from Passes 1 and 2 were combined using Samtools 1.8 (*98*) and intersected with gene annotations (Ensembl GRCh38 Release 76) supplemented with the RNY5 gene and its 10 pseudogene sequences, which were not annotated in this release, to generate the counts for individual features. Coverage of each feature was calculated by Bedtools (*99*). To avoid mis-mapping reads with embedded sncRNAs, I first intersected reads with sncRNA annotations and the remaining reads were then intersected with the annotations for protein-coding genes, lincRNAs, antisense, and other lncRNAs. To further improve the mapping rate for tRNAs and rRNAs, I combined reads mapped to tRNAs or rRNAs in the initial alignments and re-mapped them to tRNA reference sequences (Genomic tRNA Database, and UCSC genome browser website) or rRNA reference sequences (GeneBank: X12811 and U13369) using Bowtie 2 local alignment. Because similar or identical tRNAs with the same anticodon may be multiply mapped to different tRNA loci by Bowtie 2, mapped tRNA reads were combined according to their anticodon (N = 48) prior to calculating the tRNA distributions.

For miRNA analysis, sequences were mapped to miRbase using miRdeep2. Reads were normalized by dividing the numbers or reads per miRNA to the total number of miRNA reads in each sample and then this value was multiplied by one million (reads per million miRNA mapped reads – RPM).

CRISPR/Cas9 genome editing

A pX330-based plasmid expressing Venus fluorescent protein (*40*) was used to clone the gRNAs targeting Rab27a and Rab35. Two CRIPSR guide RNAs targeting each gene were selected using the CRISPR design tool (*100*). For Rab27a, gRNAs targeting exon 4 and exon 5 were selected. For Rab35 both gRNAs targeted exon 1. Both gRNAs targeting each gene were cloned simultaneously in the pX330-Venus plasmid according to the PrecisionX Multiplex gRNA Kit instructions (SBI). MDA-MB-231 cells at 25% confluency were transfected for 48h, and then cells were trypsinized and sorted for single

Venus-positive cells in a 96 well plate using a BD Influx cell sorter. Wells containing clones were allowed to expand (30 clones for Rab27a and 25 for Rab35) and knockouts for each gene were confirmed by immunoblots.

Results

Two biochemically distinct EV sub-populations are released by MDA-MB-231 cells

In order to fractionate the particulate material secreted by the highly metastatic breast cancer cell line MDA-MB-231, I developed a two-step purification strategy: differential ultracentrifugation followed by buoyant density flotation into an Optiprep linear gradient (Figure 2-1A). The Optiprep gradient showed linearity from fraction 1 to fraction 25 (Figure 2-1B). The two-step approach allowed the resolution of two distinct EV species, termed the low density (LD) and high density (HD) sub-populations (Figure 2-1C). Both EV sub-populations fractionated away from non-vesicular RNPs, as shown by the presence of Ago2 and dicer in the denser areas of the gradient (Figure 2-1C). This result is in accordance with previous studies finding the presence of Ago2 in the crude 100,000g pellet, but not associated with vesicular fractions when more stringent methods of purification were applied (*23, 40, 41*). Quantification of each protein marker along the gradient showed that classical EV reporters exhibited a differential distribution to one or the other of the EV sub-populations, whereas Ago2 and dicer did not float and remained in the denser fractions (Figure 2-1E).

The vesicular nature of both EV sub-populations was confirmed by negative electron microscopy (EM). Both EV sub-populations showed cup-shaped structures representative of vesicles as seen by EM (Figure 2-1D). Nanoparticle tracking analysis of the two EV sub-populations demonstrated that vesicle sizes largely overlapped, with HD vesicles (114.1 nm mean diameter) being slightly smaller on average than LD vesicles (121.8 nm mean diameter) (Figure 1-1F). Quantification of the total particle number of both EV sub-populations showed that there were approximately 30% more LD vesicles released into the medium in comparison to the HD counterpart (Figure 1-1G).

Altogether these results show that MDA-MB-231 cells secrete at least two distinct EV sub-populations. These two distinct EV sub-populations equilibrate to slightly different densities and fractionate away from the denser fraction consisting of vesicle-free RNPs.

The distinct EV sub-populations have different sub-cellular origins

We noticed the HD sub-population was enriched in the bona fide exosomal marker CD63 (Figure 2-1C, E). Moreover the HD sub-population showed densities ranging from 1.14-1.16 g/ml (Figure 2-1C), consistent with densities reported for exosomes (*5*). The reported density of plasma membrane of mammalian cells is lower than intracellular compartments (*101*). Thus, we hypothesize that the two distinct EV sub-populations have different sub-cellular origins.

I analyzed the total pool of proteins present in each EV sub-population by mass spectrometry. The HD sub-population was highly enriched in endosome-associated proteins whereas the LD sub-population lacked the presence of such proteins as judged

by mass spectrometry and gene ontology analysis (Figure 2-2A). I then analyzed the sub-cellular localization of two tetraspanins enriched in each EV sub-population by immunofluorescence microscopy: CD9 and CD63, enriched in LD and HD respectively (figure 2-1C). The two analyzed tetraspanins showed a distinctive sub-cellular localization, with CD9 mostly localizing in the plasma membrane (Figure 2-2B) and CD63 displaying an intracellular localization (Figure 2-2C). This result is in accordance with a recent study showing that CD9 has a mainly a plasma membrane sub-cellular localization, whereas CD63 localizes to endosomes (*102*).

At least three Rab-GTPase proteins have been shown to play a role in the fusion of multivesicular bodies with the plasma membrane thus promoting exosomal release (*17-19*). We hypothesize that knocking down these proteins would result in the impairment of the secretion of bona fide exosomes whereas secretion of vesicles derived from the plasma membrane may remain unchanged. Thus I systematically knocked downed Rab11, Rab27a and Rab35 using CRISPR interference-CRISPRi (Figure 2-2D). Extracellular vesicles were isolated from conditioned medium of WT and knock-downed cells and analyzed by immunoblots for the levels of two tetranspanins defining the two EV sub-populations: CD63 as a protein marker defining the HD sub-population and CD9 defining the LD sub-population (Figure 2-1C). EVs derived from Rab27a and Rab35 knock downed cells reported a slight difference in the secretion of CD63-positive vesicles, but not in CD9-positive vesicles (Figure 2-2E), thus we decided to investigate their role further.

The Rab27a protein has been reported to be involved in the fusion of MVBs to the plasma membrane in different cell lines including dendritic cells (*103*), cervical cancer cells (*18*), melanoma cells (*62*), bladder cancer cells (*85*), lung cancer cells (*104*) and breast cancer cell lines (*105, 106*). We created a Rab27a knock-out cell line using CRISPR-Cas9 technology (Figure 2-2F). Extracellular vesicles were isolated from conditioned medium of WT and Rab27a knock-out cells and analyzed by immunoblot for the levels of CD63 and CD9. EVs derived from the Rab27a KO background showed a dramatic decrease of CD63 signal, while secreting seemingly similar levels of CD9-positive vesicles (Figure 2-2G) in comparison to the WT levels. Our results are in accordance with a previous study showing that the secretion of CD63-positive, but not of CD9-positive vesicles, is Rab27a dependent (*91*). However, despite the dramatic reduction of secretion of CD63 positive vesicles under the Rab27a KO background, there was only a subtle reduction for the HD enriched marker flotillin-2 (Figure 2-2H). Thus suggesting that the HD vesicle fraction may comprise at least two different types of EV sub-populations, with at least one of them having its origin in the endocytic pathway. This possibility is consistent with our previous demonstration that EVs secreted by HEK293 cells contain a population of vesicles that bind to immobilized CD63 antibody and a fraction containing flotillin-2 that does not (*40*).

The Rab35 protein has being implicated in mediating exosomal secretion in oligodendrocytes (*19*). We created a Rab35 knock out cell line (Figure 2-2F). When EVs derived from the Rab35 KO background were probed for CD63 and CD9 signals, no apparent difference was observed in comparison to EVs derived from WT cells (Figure 2-

2G). Thus, Rab27a but not Rab35, is involved in the secretion of CD63-positive vesicles derived from MDA-MB-231 cells. Collectively, our data are most consistent with HD vesicles having at least one type of vesicle originating in the endocytic pathway, whereas LD vesicles may bud directly from the plasma membrane.

Characterization of the RNA content of the distinct EV sub-populations

Since RNA has been proposed to serve a role in EV mediated intercellular communication (*27, 30-32, 37-39, 55, 56*) we decided to investigate the RNA cargo of the distinct EV sub-populations. An initial qualitative assessment of RNA profiles using a Bioanalyzer showed that both LD and HD sub-populations share a predominant peak at ~70 nucleotides (Figure 2-3A). We decided to use thermostable group II intron reverse transcriptase sequencing (TGIRT-seq) to investigate the EV RNA cargo further. The TGIRT enzyme is a thermostable, highly processive reverse transcriptase with a proficient template-switching activity used for RNA-seq adapter addition (*107*). This enzyme's properties allowed us to generate comprehensive cDNA libraries that included highly structured and modified transcripts. Also due to its properties, TGIRT allows the generation of cDNA libraries from low input material (*26, 108*). The data obtained from TGIRT-seq was normalized to total number of small non-coding RNA reads per sample. In agreement with our previous findings (*26*), I found that the most abundant small non-coding transcripts in EVs are tRNAs (representing 88.6% of HD and 96.3% of LD transcripts – Figure 2-3B). Interestingly, I found that among all the tRNA isotypes, both tRNA lysine isoacceptors are the most abundant transcripts (Figure 2-3D). Moreover, both tRNA lysine isoacceptors are highly enriched in both EV sub-populations in comparison to intracellular levels (Figure 2-3D). Furthermore, I found that most of the tRNA reads in both EV sub-populations start at approximately position 16 (Figure 2-3C) which may represent a reverse transcription stop at the same unidentified EV-enriched D-loop modification described previously (*26*). Thus, both EV sub-populations share a modified tRNA version as their most abundant small non-coding transcript.

Distinct molecular mechanisms of miRNA sorting govern the discrete small extracellular vesicle sub-populations

The initial Bioanalyzer analysis showed an apparent enrichment of transcripts smaller than 70 nucleotides in the HD sub-population (Figure 2-3A) that I hypothesized represented miRNAs. As such, I considered the possibility that the distinct EV sub-populations might differ in their miRNA content. To further analyze the miRNA content of the HD and LD vesicles, I used two orthogonal approaches: targeted miRNA profiling and an unbiased TGIRT - sequencing approach. Using the commercially available Firefly particle technology, Discovery Panel (Abcam), I compared the abundance of 408 miRNAs in the LD and HD EV sub-populations. Out of the ~400 tested miRNAs, 312 showed a detectable signal in the HD sub-population and 217 were detected in the LD counterpart. I found 21 miRNAs present in the HD vesicles that were enriched at least 10-fold in comparison to their intracellular levels (Figure 2-4A). In contrast, only 2 miRNAs detected in the LD sub-population showed over a 10-fold change relative to cell lysate (Figure 2-4B). Both of the apparent LD-enriched miRNAs were also present as highly enriched in the HD sub-population (Figure 2-4A). Their enrichment in the LD sub-population was much lower than in the HD sub-population (Figure 2-4E), suggesting

their detection in the LD subpopulation may have been the result of small amounts of HD vesicles co-fractionating with the LD vesicles during flotation.

I then analyzed the TGIRT sequencing data generated previously, focusing specifically on miRNAs. I found that miRNAs represented a relatively low proportion of total reads (3.4% and 1% of the mapped reads for HD and LD, respectively), resulting in detection of fewer miRNAs by TGIRT-seq than by Firefly profiling. Nevertheless, TGIRT-seq recapitulated the general trends observed by Firefly: multiple miRNAs enriched by at least 10-fold in the HD sample compared to cells (Figure 2-4C) and fewer enriched miRNAs in the LD sample compared to cells (Figure 2-4D). Out of the 4 apparent LD enriched miRNAs detected by TGIRT-seq, 3 were also detected as highly enriched in the HD sub-population, suggesting again that their presence was a result of carryover material overlapping the buoyant density gradient fractions. The total number of uniquely detected HD miRNAs by TGIRT-seq (detectable in the HD sample, but absent in the intracellular lysate) was 21. Comparatively, only 5 miRNAs were uniquely found in the LD sub-population (Figure 2-4F).

I next validated the top EV-enriched/unique miRNAs per vesicle sub-population using RT-qPCR. Relative levels for each EV sub-population were compared to intracellular levels. I found that all of our top HD candidates were validated by this method. As expected, the tested HD miRNAs were highly enriched in the HD EV sample relative to the intracellular levels. The HD/intracellular lysate ratio ranged from ~50-fold to ~1,000-fold (Figure 2-4G). In contrast, none of our top LD miRNA candidates were validated as being highly enriched in comparison to their intracellular levels. The LD/intracellular lysate ratio ranged from ~0.5-fold to ~5-fold (Figure 2-4H). I then compared relative levels for each EV sub-population to each other in order to test for miRNA vesicle specificity. HD-enriched miRNAs were highly depleted in the LD vesicles, as seen by >10-fold increase in their HD/LD ratio (Figure 2-4G). In contrast, the tested LD miRNAs had roughly the same abundance or were depleted in comparison to their relative levels in HD vesicles (Figure 2-4H). This proved to be true not only for the few miRNAs tested by RT-qPCR, but was also a general trend for multiple miRNAs detected by Firefly profiling, which demonstrated that many miRNAs were more abundant in the HD vesicles than in their LD counterpart (Figure 2-4I).

These results suggest that a selective miRNA sorting mechanism into EVs occurs in the biogenesis of the HD sub-population but less so or not at all in the LD sub-population. Thus, we identified two distinct EV sub-populations that are released by MDA-MB-231 cells that utilize fundamentally different miRNA sorting mechanisms (selective vs. non-selective).

MiR-122 is a bona fide secreted HD resident miRNA

In order to confirm that the HD-enriched miRNAs were bona fide vesicle residents and not merely externally EV associated I performed RNase protection assays. The high-speed pellet fraction was exposed to ribonuclease in the absence or presence of non-ionic detergent, protected RNA was quantified by RT-qPCR. Highly enriched (>50-fold change HD/cellular lysate) and specific (>20-fold change HD/LD) HD miRNAs were

selected for this assay. All of the highly enriched, specific HD miRNAs tested proved to be EV encapsulated: they were resistant to RNase in the absence but not in the presence of detergent (Figure 2-5A).

I selected miR-122 for further analysis as it has been shown that circulating miR-122 can serve as a prognostic biomarker for metastasis in patients with breast cancer (*109*) and very specifically that MDA-MB-231 cell-derived exosomal miR-122 may promote metastasis by reprogramming glucose metabolism in the premetastatic niche (*32*). Because miR-122 is present in the fetal bovine serum (FBS) used to culture cells (*110*), I employed clarified FBS depleted of sedimentable particles to confirm that miR-122 was released by MDA-MB-231 cells. I observed an accumulation of miR-122 over time in conditioned medium exposed to MDA-MB-231 cells, whereas the level of miR-122 slightly decreased over time when conditioned medium was incubated in the absence of cells (Figure 2-5B). This suggested that newly synthesized MDA-MB-231 cell-derived miR-122 was secreted into the extracellular space. Moreover, when the levels of mir-122 were tested across the iodixanol linear gradient, I observed co-fractionation of mir-122 with the HD fraction, substantially separated from RNP particles at the bottom of the gradient (Figure 2-5C, D, E). Thus I confirmed that MDA-MB-231 cells actively secrete miR-122. Moreover the vast majority of secreted miR-122 resides inside HD vesicles.

Discussion

Here I document the fractionation of secreted material sedimented at high-speed ultracentrifugation derived from MDA-MB-231 cells. I describe the purification of two distinct EV-subpopulations with different sub-cellular origins each of them employing different mechanisms of miRNA sorting (Figure 2-6).

Different molecular mechanisms of miRNA sorting

By using differential ultracentrifugation followed by density gradient flotation, I document the separation of two EV populations, likely of different sub-cellular origin, which appear to employ different molecular mechanisms of miRNA sorting. I found that EVs with a density ranging from 1.09-1.11 g/ml (LD) may originate in the plasma membrane, whereas the EVs with a density ranging from 1.14 to 1.16 g/ml (HD) are more characteristic of exosomes, originating in the endocytic pathway. Some miRNAs are non-selectively sorted into both EV sub-populations, as shown by their similar relative abundances in the cellular lysate and both EV sub-populations (Figure 2-4H). I observed a slight enrichment for non-selectively sorted miRNAs in the HD sub-population (Figure 2-4A) and a minor depletion of non-selectively sorted miRNAs in the LD sub-population (Figure 2-4B).

Non-selective sorting may correlate to the fraction of miRNAs unbound to the RNA-induced silencing complex (RISC). Association of miRNAs with RISC inside cells varies greatly: >1,800-fold for different miRNAs present in the same cell line (*111*). As a result, some miRNAs may exist in the cytoplasm as mostly RISC-free species. Our data show that Ago2 and Dicer are not present in EVs derived from MDA-MB-231 cells (Figure 2-1C, E). Thus, we suggest that the non-selectively sorted miRNAs may

represent the cytosolic pool of RISC-free miRNAs, not actively engaged in silencing, and their presence in EVs is the result of a passive mechanism of sorting.

P-bodies have been shown to be closely associated with MVBs (*112*) and the levels of target mRNAs can influence the levels of miRNA association with RISC (*111*), as well as miRNA secretion as detected in the high-speed pellet fraction sedimented from conditioned medium (*80*). Thus, miRNAs released from P bodies and freed from RISC may be proximal to endosomes engaged in the production of HD vesicles and distant from the site of LD vesicle biogenesis, thus resulting in a slight enrichment and slight depletion of these miRNAs in the HD and LD vesicles, respectively.

I focused on the set of highly enriched HD miRNAs for a closer investigation, since these miRNAs seem to undergo a highly selective and active secretion. These selectively sorted miRNAs may belong to the pool of pre-existing RISC-free miRNAs, or may be displaced from RISC prior to or concomitant with sorting into HD vesicles. Importantly, I found that this mechanism of selective miRNA sorting is specific for the HD sub-population. I speculate that the sorting machinery required to enrich certain miRNAs in exosomes is specific for MVBs, and it is absent at the site of origin of the LD vesicles. However, enveloped viruses that bud from the cell surface clearly do sort their RNA genomes in a selective manner (*113, 114*) and the same may be true of other RNA biotypes in the LD sub-population. We previously showed that tRNAs with an unknown post-transcriptional modification in the D-loop are highly enriched in EVs in comparison to the cells from the same culture (*26*). Similarly, I detected apparently similarly modified tRNA species in both EV sub-populations secreted from MDA-MB-231 cells (Figure 2-3C). Interestingly I also found that both tRNA lysine isoacceptors are highly enriched in both EV sub-populations in comparison to cellular lysates (Figure 2-3D). Both tRNA lysine isoacceptors are selectively packaged in HIV virions (*115*), but only tRNA lysine UUU is annealed to the HIV genome. This suggests there might shared cellular machinery responsible to package tRNA lysine (specially tRNA lysine CUU) in HIV virions and EVs. The machinery for the active sorting of the modified tRNA/tRNA lysine may be shared for both EV sub-populations.

It is widely suggested that EV miRNAs play a role in intercellular communication (*30-32, 34-39, 63*). Thus, purifying EV sub-populations, with differential miRNA content, is critical in order to accurately assess their role in intercellular communication. One study purified different EVs based on their buoyant densities and showed that the LD sub-population triggered the differential expression of 257 genes in recipient cells, whereas the HD sub-population caused the differential expression of 1,116 genes (*22*). Although this study did not determine if any gene expression differences were the result of miRNA transfer, the effect could be partially explained by the distinct miRNA composition of the two EV sub-populations. Further studies using purified EV sub-populations are needed to understand the physiological consequence of selective vs. non-selective sorting mechanisms. Additionally, it needs to be determined if the effects are directly attributable to miRNAs or some other constituents of the two EV populations.

Significance of selective miRNA sorting into EVs in cancer

Intercellular communication mediated by miRNAs is a subject of considerable interest, but in most cases, it remains unproven. Many studies using cancer cells suggest EV-mediated miRNA transfer to recipient cells (*30-32, 34-39, 63*), possibly to promote metastasis. In the context of breast cancer for example, EV resident miR-122 has been suggested to promote metastasis through reprogramming glucose metabolism in the resident cells of the premetastatic niche (*32*). In another example, EV-associated miR-105 may promote invasiveness in metastatic breast cancer by disruption of tight junctions, thus promoting vascular permeability and tumor cell invasion (*34*). For EV miRNAs to be functionally transferred to recipient cells, fusion of the EV membrane with a recipient membrane (either plasma membrane or endosomal membrane) must precede the presentation of an exogenous miRNA to the target cell cytoplasm. I find that the highly-selected HD miRNAs in exosomes are not bound to Ago2 (Figure 2-1C, E and Figure 2-5D, E). Thus, the usual function of a miRNA would require loading from the EV interior to a cytoplasmic RISC. Unfortunately, nothing is known about the cytoplasmic presentation of miRNAs derived from EVs.

Alternatively, tumor cell invasiveness may depend on the selective secretion and disposal of tumor suppressor miRNAs. Such a suggestion has been made for miR-23b (*85*) and miR-100 (*116*), which are secreted at high efficiency by cancer cells. These two mechanisms, intercellular communication and disposal, are not mutually exclusive. It is possible that EVs containing detrimental miRNAs have been repurposed to serve as shuttles for intercellular communication, as has been suggested for the secretion of miR-142 (*33*). Moreover another highly selected miRNA reported here, miR-122 has been suggested as a tumor suppressor with cells developing greater invasiveness on loss of this miRNA (*117, 118*). Nonetheless, the importance of EV resident miR-122 as a molecule to convey a message to promote metastasis has also been suggested (*32*).

A bigger challenge in the field of cancer and EV biology is that claims for a role of EV-associated miRNAs are based on crude preparations of sedimentable particles or co-culture experiments. Here I show that Ago2 is secreted to the extracellular space as a RNP, not vesicle-associated, that co-isolates with EVs in the high-speed pellet (Figure 2-1C, E). Thus, clear demonstration that the implicated miRNAs in cancer development are enclosed within a membrane vesicle is still missing. Studies purifying, characterizing, understanding the cellular biology behind different EV sub-populations and their functional effects in the context of cancer will prove important in the future.

Figures

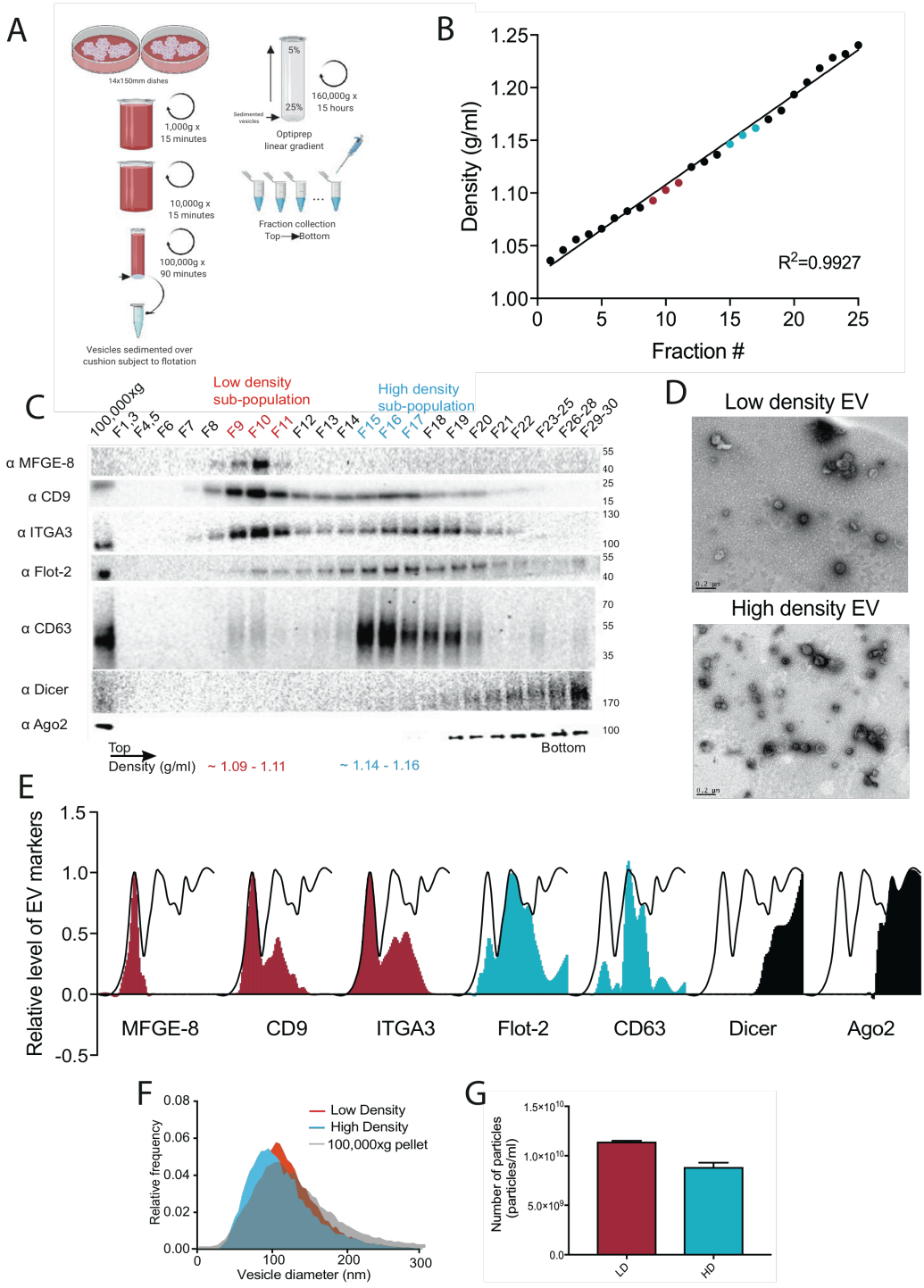


Figure 2-1: Two biochemically distinct EV sub-populations are released by MDA-MB-231 cells.

(A) Schematic showing the two-step purification methodology. Differential ultracentrifugation was followed by buoyant density flotation in a linear iodixanol gradient. (B) Linearity of iodixanol density gradient. Calculated iodixanol fraction

densities collected from the top to bottom of the gradient (fractions 1 to 25) as measured by refractometry. In red, fractions corresponding to the LD sub-population. In blue, fractions corresponding to the HD subpopulation (C) Immunoblot across the iodixanol gradient for classical EV markers as well as other non-vesicular RNP components. The two discrete subpopulations are indicated. CD63, a glycosylated protein, migrates heterogeneously. (D) Representative electron micrographs of negative stained HD and LD samples. (E) Quantification of individual proteins as in C. The black line delineates the maximum signal across the gradient showing three distinctive areas. The first, second and third peak represent the LD, HD and nonvesicular RNP fractions. (F) Nanoparticle tracking analysis showing the size distribution of the HD and LD sub-populations. The high-speed pellet is also shown. (G) Quantification of the particle number per EV sub-population using Nanoparticle tracking analysis. Data plotted are from two independent experiments; error bars represent standard deviation from independent samples.

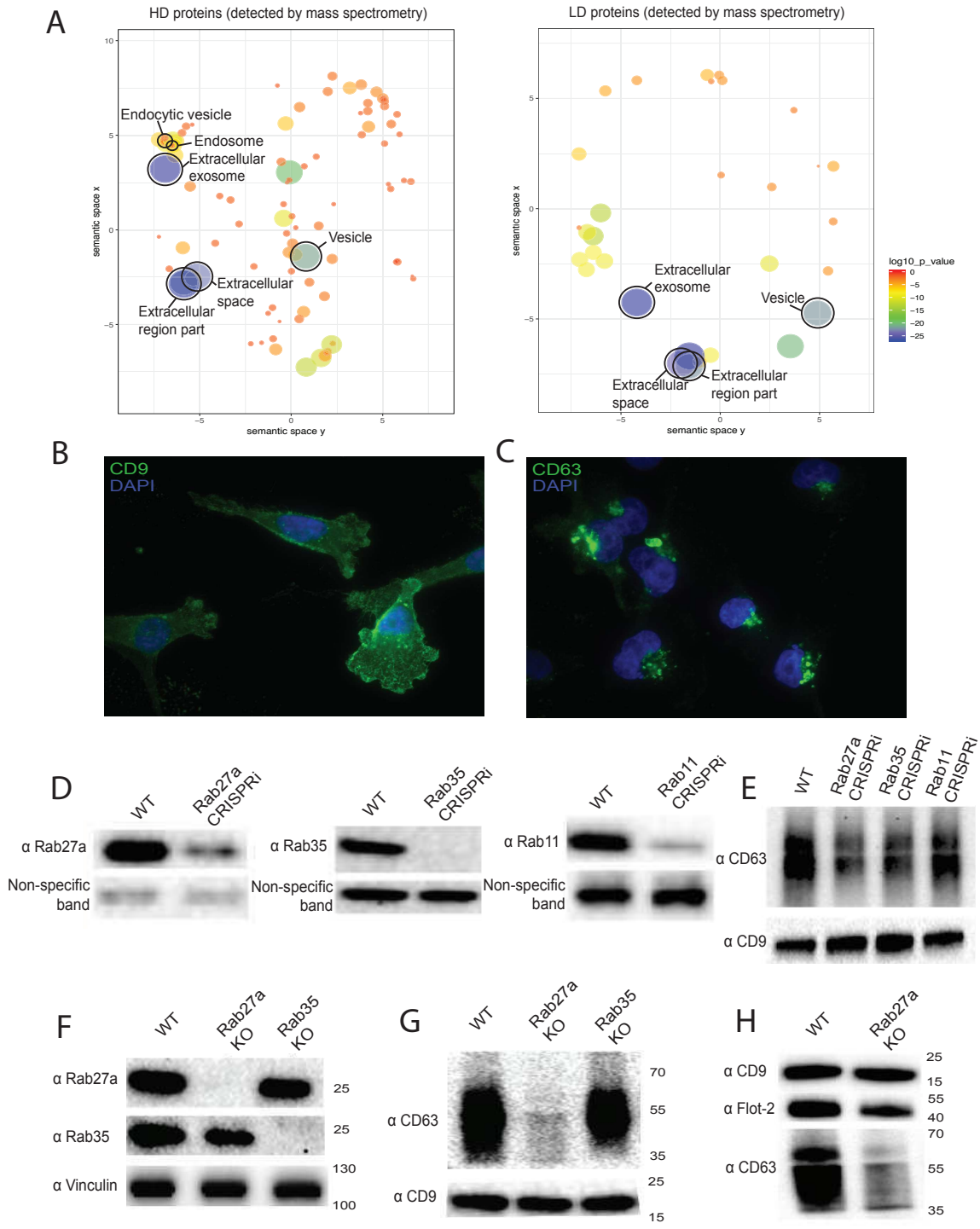


Figure 2-2: The two biochemically distinct EV sub-populations co-fractionate with membranes of different sub-cellular origin.

(A) Gene ontology analysis for sub-cellular localization of proteins that coincide with HD and LD proteins detected by mass spectrometry. The gene ontology analysis compares the list of proteins given per group to the human genome frequency. The detection of enrichment for a certain organelle/ localization is represented as circles, with

the size of circles correlating to log10 p-value. (B, C) Immunofluorescence for two tetraspanins enriched in each EV sub-population. CD9 and CD63 enriched in LD and HD respectively. (D) Analysis of depletion of Rab proteins by CRISPRi. Immunoblots for intracellular levels of Rab11, Rab27a and Rab35 upon CRISPRi depletion are shown. Immunoblots were normalized per total protein concentration. Non-specific bands for each immunoblot are shown as evidence of equal amount of proteins loaded per lane. (E) Immunoblots for CD63 and CD9 in EVs secreted by WT, Rab11, Rab27a and Rab35 depleted cells are shown. The amounts of loaded EVs were normalized by total cell number. (F) Analysis of WT, CRISPR-Cas9 Rab27a KO and CRISPR-Cas9 Rab35 KO by immunoblot. Rab27a, Rab35 and vinculin are shown. (G) Immunoblots for CD63 and CD9 in EVs secreted by WT, Rab27a KO cells and Rab35 KO cells. The amounts of loaded EVs were normalized by total cell number. (H) Immunoblots for CD63, CD9 and flotillin-2 in EVs secreted by WT and Rab27a KO cells. The amounts of loaded EVs were normalized by total cell number.

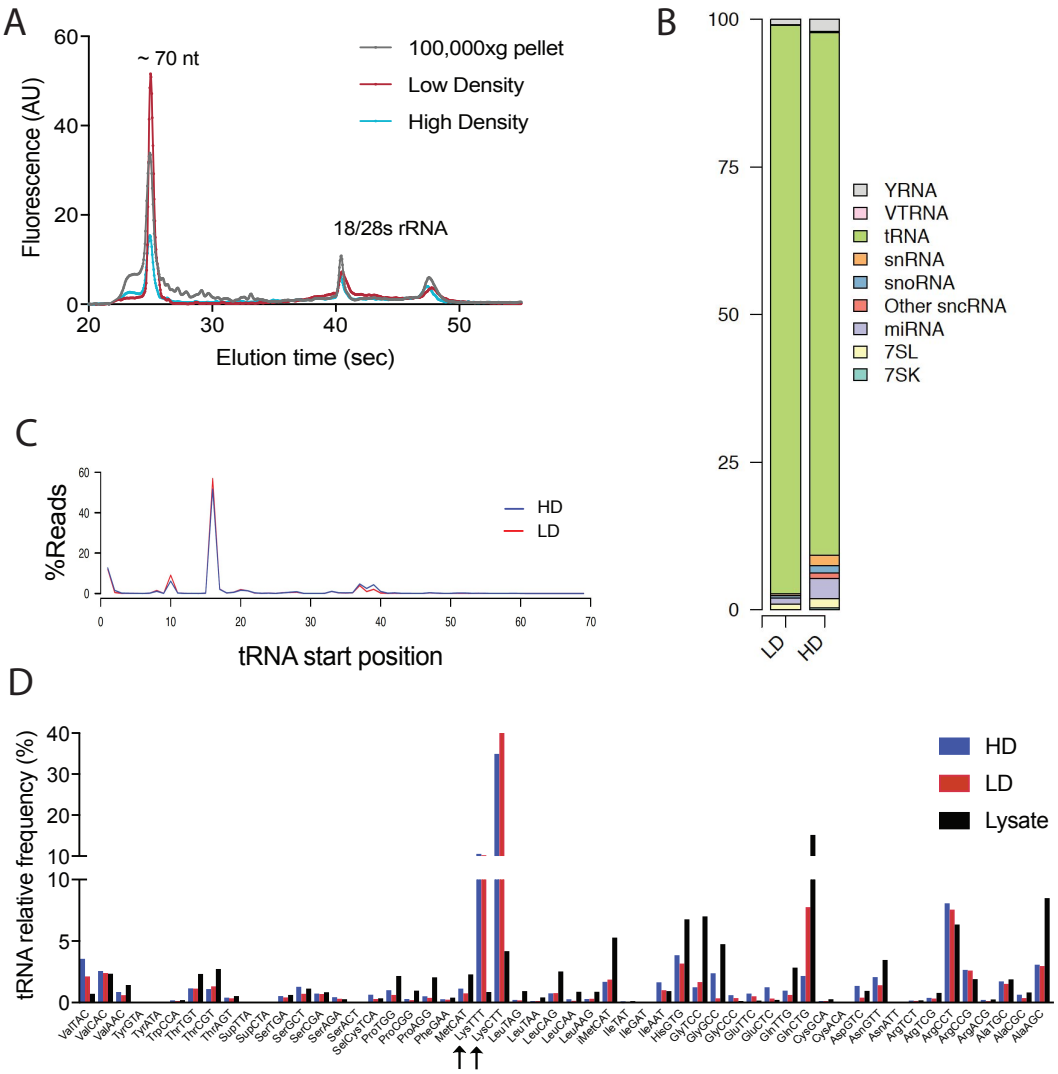
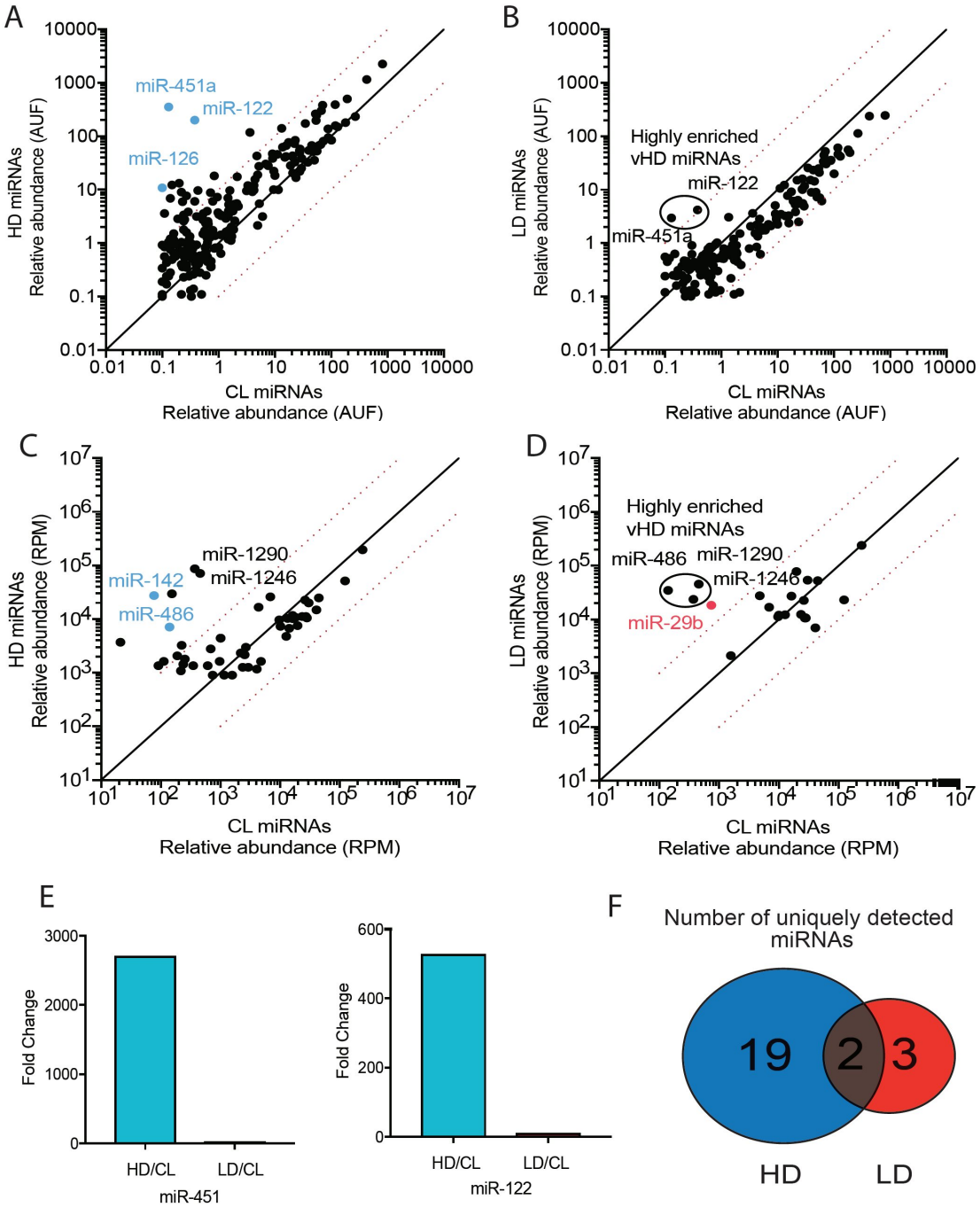


Figure 2-3: tRNAs are the most abundant transcript in both EV sub-populations.

(A) Bioanalyzer analysis of the HD and LD RNA. The high-speed pellet RNA is also shown. A prominent ~70nt peak is observed. (B) Bar graph showing the distribution of all small non-coding RNA transcripts found in HD and LD sub-populations. tRNAs represent the majority of transcripts present in both EV sub-populations (C) tRNA start site distribution showing the majority of tRNA reads start at position 16, previously described as the site of an EV-specific D-loop modification (*26*). The putative D-loop modification is present at relatively similar levels in both the HD and LD sub-populations. (D) tRNA isotypes frequency. The frequency of each tRNA isotype was calculated in comparison to the total number of tRNA counts per sample. Both tRNA lysine isoacceptors are enriched in HD and LD sub-populations in comparison to the intracellular levels. tRNA lysine isoacceptors are highlighted with arrows.


A
HD miRNAs
Relative abundance (AUF)
CL miRNAs
Relative abundance (AUF)
miR-451a
miR-122
miR-126
B
LD miRNAs
Relative abundance (AUF)
CL miRNAs
Relative abundance (AUF)
Highly enriched
vHD miRNAs
miR-122
miR-451a
C
HD miRNAs
Relative abundance (RPM)
CL miRNAs
Relative abundance (RPM)
miR-1290
miR-1246
miR-142
miR-486
D
LD miRNAs
Relative abundance (RPM)
CL miRNAs
Relative abundance (RPM)
Highly enriched
vHD miRNAs
miR-486
miR-1290
miR-1246
miR-29b
E
Fold Change
HD/CL
LD/CL
miR-451
Fold Change
HD/CL
LD/CL
miR-122
F
Number of uniquely detected
miRNAs
19
2
3
HD
LD

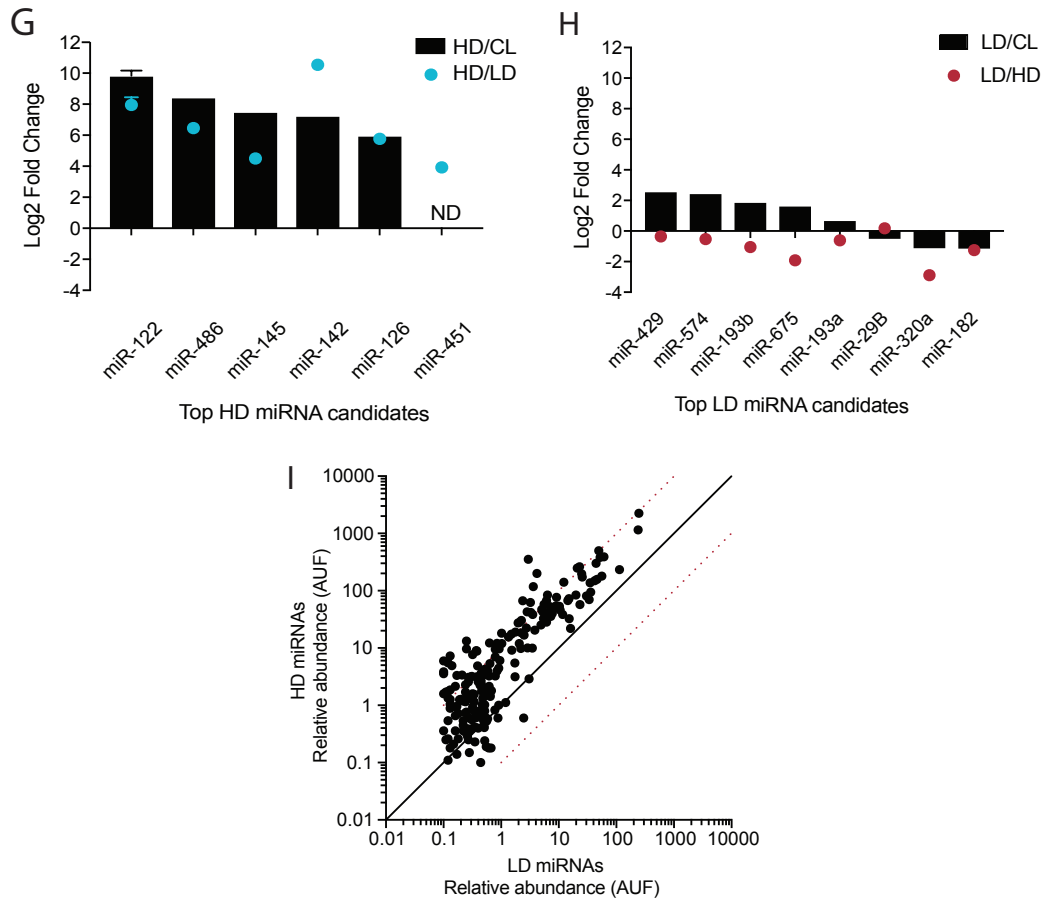


Figure 2-4: Distinct molecular mechanisms of miRNA sorting govern the discrete small extracellular vesicle sub-populations.

(A, B) MicroRNA profiling of HD and LD EV sub-populations. Scatterplots showing the relative abundance (as arbitrary units of fluorescence (AUF)) for miRNAs detected in HD and LD sub-populations relative to cellular levels, as detected by Firefly profiling and normalized per ng of total RNA. In blue, HD miRNAs that were selected for further validation by RT-qPCR. Circled miRNAs in LD represent those also found to be enriched in the HD-subpopulation. (C, D) TGIRT-sequencing of HD and LD EV sub-populations. Scatterplots showing the relative abundance (as reads per million (RPM)) for miRNAs detected in HD and LD sub-populations relative to CL levels. Normalized per total number of miRNA reads. In blue, HD miRNAs that were selected for further validation by RT-qPCR. In red, LD miRNAs selected for further validation by RT-qPCR. Circled miRNAs in LD represent those also found to be enriched in the HD-subpopulation. (E) Circled LD miRNAs from (B) are highly enriched in the HD sub-population. Fold difference HD/CL or LD/CL for miR-451a and miR-122 in linear scale are shown. Data extracted from Firefly profiling. (F) Venn Diagram showing the total number of HD and LD uniquely found miRNAs (miRNAs with non-detectable signal in cellular lysates) detected by TGIRT-seq. (G, H) RT-qPCR validation of selected miRNAs. Log2 fold change for top HD and LD candidate miRNAs in HD or LD relative to cells (HD/CL or LD/CL, respectively) and HD relative to LD or vice versa (HD/LD or

LD/HD, respectively). MiRNA species were quantified by RT-qPCR and normalized per ng of total RNA. Data plotted for miR-122 are from three independent experiments (biological replicates), each independent experiment with triplicate qPCR reactions; error bars represent standard deviation from independent samples. (I) Scatterplots showing relative abundance of miRNAs (AUF) detected in HD and LD sub-populations normalized per ng of total RNA, as detected by Firefly profiling. CL: cellular lysate. Dashed red lines in A, B, D and E, represent 10-fold differences.

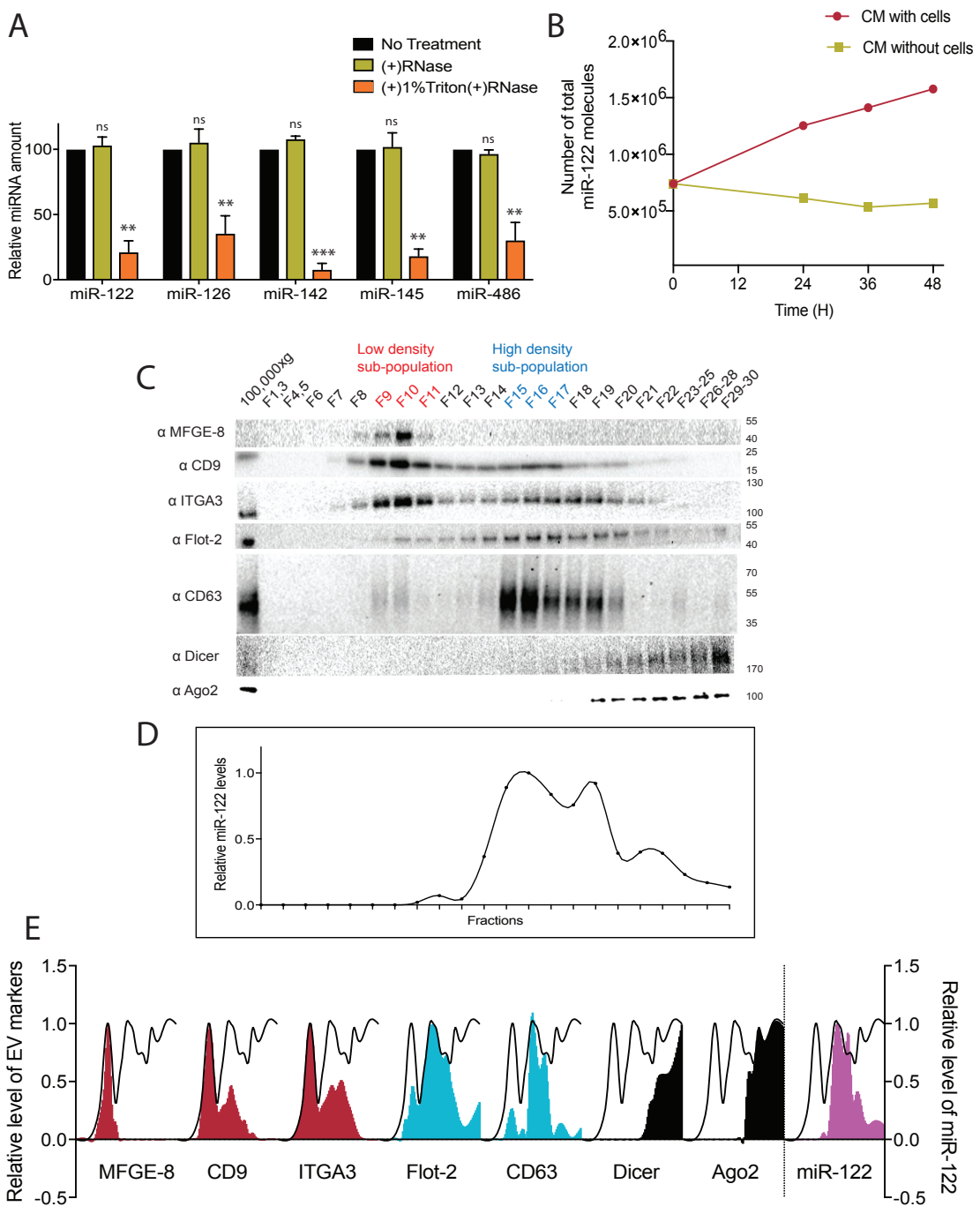


Figure 2-5: miR-122 is a bona fide secreted HD resident miRNA.

(A) RNase protection of highly enriched HD miRNAs quantified by RT-qPCR. Isolated EVs were treated with or without RNase If and or Triton X-100. Data plotted are from two independent experiments, each independent experiment with triplicate qPCR reactions; error bars represent standard deviation from independent samples. (B) MiR-122 accumulates in conditioned medium (CM) over time. Quantification of total number of miR-122 molecules in CM. CM was incubated with or without MDA-MB-231 cells for a period of 48 h. Samples were collected at the indicated time points. Cell debris and

apoptotic bodies were removed by low- and medium-speed centrifugation as described in methods. (C) Iodixanol linear gradient showing the distribution of EV and non-vesicular RNP markers. (D) Quantification of miR-122 by RT-qPCR across the linear gradient. The majority of MiR-122 co-fractionates with the HD subpopulation. (E) Quantification of individual proteins as in (C) and miR-122 as in (D) across the gradient. The black line delineates the maximum signal across the gradient showing the three distinct components. The first, second and third peak represent the LD, HD and non-vesicular RNP fractions. Statistical significance was performed using Student's t-test (**$p<0.01$, ***$p<0.001$, ns = not significant).

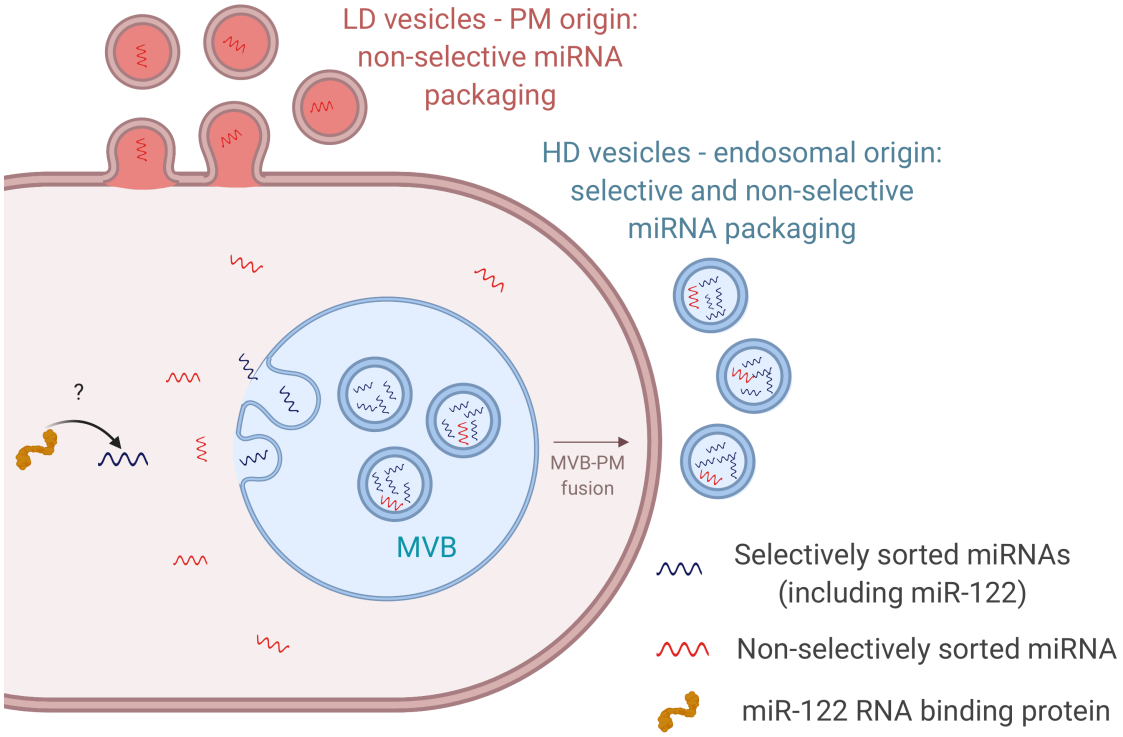


Figure 2-6: Diagram representing the current model of EV secretion by MDA-MB-231 cells and their distinct mechanisms of miRNA sorting.

At least one vesicular species within the HD sub-population has its origin in the endocytic pathway, representing the classical exosomes. We propose a selective mechanism of miRNA sorting occurring at the site of HD vesicle biogenesis where miR-122, along with its cognate RNA binding protein, is targeted for capture into a bud invaginating into the interior of an endosome. The discovery of the miR-122 binding protein required for this process is described in *Chapter 3*. LD vesicles may originate at the plasma membrane, representing shedding vesicles. There may be no selective miRNA sorting occurring at the site of LD vesicle biogenesis. Non-selective miRNA sorting occurs in both the HD and LD sub-populations. MVB: multivesicular body, PM: plasma membrane.

Chapter 3: The Lupus La protein is required for selective miRNA sorting into EVs derived from MDA-MB-231 cells

Introduction

The presence of miRNAs in EVs has been widely accepted (*119*). In the past few years much attention has been given to EV miRNAs in the context of cancer biology (*120*). It is proposed that cancer cells secrete vesicular miRNAs either as a means of intercellular communication, changing the gene expression of recipient cells (*30-32, 34-39, 63*), or as a way to dispose of tumor-suppressor miRNAs (*33, 85, 116*). Despite the potential physiological implications of EV miRNA in cancer biology very little is known about the molecular mechanisms by which they are sorted. In *Chapter 2,* I described the discovery of two distinct EV sub-populations produced by MDA-MB-231 cells with differential mechanisms of miRNA sorting: selective and non-selective. In this chapter, I present studies on the proteins involved in mediating the selective process of miRNA sorting into EVs.

By using classical cellular fractionation followed by an *in vitro* miRNA packaging assay I found potential RNA-binding proteins involved in mediating miR-122 sorting. A combination of genetic and biochemical tools then allowed me to show that the La protein is required for the secretion of selectively sorted miRNAs into EVs formed *in vitro* and *in vivo*.

Materials and Methods

Cell lines, media and general chemicals

MDA-MB-231 cells were cultured in DMEM with 10% FBS (Thermo Fisher Scientific, Waltham, MA). MDA-MB-231 cells were confirmed by short tandem repeat profiling (STR) and tested negative for mycoplasma contamination. For EV production, cells were seeded at ~10% confluency in 10 % exosome-depleted FBS (System Biosciences, Palo Alto, CA) in 150mm CellBIND tissue culture dishes (Corning, Corning, NY) containing 30ml of growth medium. EVs were collected when cells reached ~80% confluency (~72h). Unless otherwise noted, all chemicals were purchased from Sigma Aldrich (St. Louis, MO).

Extracellular vesicle purification

Conditioned medium (720 ml for TGIRT-seq and miRNA profiling and 420 ml for all other experiments) was harvested from 80% confluent MDA-MB-231 cultured cells. All subsequent manipulations were performed at 4°C. Cells and large debris were removed by centrifugation in a Sorvall R6+ centrifuge (Thermo Fisher Scientific) at 1,000xg for 15 min followed by 10,000xg for 15 min in 500 ml vessels using a fixed angle FIBERlite F14-6X500y rotor (Thermo Fisher Scientific). The supernatant fraction was then centrifuged at ~100,000xg (28,000 RPM) onto a 60% sucrose cushion in buffer C (10mM HEPES pH 7,4, 0.85% w/v NaCl) for 1.5 h using two SW-28 rotors. The interface over the sucrose cushion was collected and pooled for an additional ~130,000xg

(32,500 RPM) centrifugation over a 60% sucrose cushion in a SW41 rotor for 15 h. The collected interface from the first sucrose cushion should not exceed a sucrose concentration of 21%, as measured by refractometry, for the second centrifugation in the SW41 to be successful. Higher concentrations of sucrose cause the EVs to equilibrate at the ambient buoyant density, impeding sedimentation. For purification of EV sub-populations based on their distinct buoyant density, the cushion-sedimented vesicles were collected and mixed with 60% sucrose to a final volume of 4ml (sucrose final concentration ~40%). Layers of 1.5 ml of 25%, 20%, 15%, 10% and 5% iodixanol (Optiprep) solutions in buffer C were sequentially overlaid and samples were centrifuged at ~160,000xg (36,000 RPM) for 15 h in a SW41 rotor. Fractions (400ul) from top to bottom were then collected and mixed with Laemmli buffer for immunoblot analysis or RNA was extracted using a mirVana miRNA isolation kit (Thermo Fischer Scientific). In some cases, such as the detection of CD63, La or nucleolin, the floated fractions were centrifuged in a SW41 rotor at ~130,000g in order to concentrate them and improve their detection. This method produced a linear density gradient from 1.036 to 1.24 g/ml.

For EV purification in bulk (without discriminating among EV sub-populations), the EVs collected after the first SW41 centrifugation were mixed with 60% sucrose to a final volume of 9 ml (sucrose concentration ~50%). Layers of 1.5 ml of 40% and 10% sucrose were overlaid and the sample was centrifuged at ~160,000xg (36,000 RPM) for 15 h in a SW41 rotor. Buoyant vesicles spanning from the 40% sucrose region and the 10/40% interface were collected and mixed with Laemmli buffer or RNA was extracted with a mirVana miRNA isolation kit. For further processing to improve protein detection, buoyant vesicles were first diluted in buffer C (10mM HEPES pH 7,4, 0.85% w/v NaCl) to a final volume of 5 ml and soluble content was released by adding 10% Triton X-100 (TX-100) to a final concentration of 1%. Samples were then homogenized by the use of a vortex mixer for 1 min, cooled on ice for 10 min and homogenized again for an extra min. To precipitate proteins, I added 4 volumes of cold acetone (previously stored at -20°C) and the mixture was incubated at -20°C overnight. Precipitated proteins were collected using a Sorvall RC 6+ centrifuge by centrifugation at 16,000xg (11,600 RPM) for 30 min at 4°C in a FIBERlite F211-8X50Y rotor (Thermo Fischer Scientific). The pellet fraction was then resuspended in Laemmli buffer for analysis by immunoblotting.

For proteinase K protection assays, the EVs were collected by centrifugation at ~100,000xg (28,000 RPM) for 1.5 h using two SW-28 rotors. Pellet fractions resuspended in buffer C were pooled and centrifuged at ~160,000xg (36,000 RPM) in a SW55 rotor for 1 h. The pellet was resuspended in buffer C and split into 3 equal aliquots. One sample was left untreated, another sample treated with 5ug/ml proteinase K on ice for 20 min, and the last was mixed with TX-100 to a final concentration of 1% prior to proteinase K treatment. Proteinase K was inactivated with 5mM phenylmethane sulfonyl fluoride (PMSF) (5 min on ice) and samples were then mixed with Laemmli buffer for immunoblot analysis.

Immunoblots

Exosomes were prepared by mixing sedimented vesicles with 1X Laemmli buffer,

or 1X Laemmli buffer without DTT when detection of CD63 was performed. Cell lysates were prepared by adding lysis buffer (10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 0.1% sodium dodecyl sulfate, 0.5% sodium deoxycholate, 1% Triton X-100, 10% glycerol) to cell pellets. Protein was quantified using a BCA Protein Assay Kit (Thermo Fischer Scientific), and the selected amount was mixed with Laemmli buffer. Samples were heated at 95°C for 5 min and separated on 4-20% acrylamide Tris-Glycine gradient gels (Life Technologies). Proteins were transferred to polyvinylidene difluoride membranes (EMD Millipore, Darmstadt, Germany), blocked with 5% bovine serum albumin in TBST and incubated overnight with primary antibodies. Blots were then washed with TBST, incubated with anti-rabbit or anti-mouse secondary antibodies (GE Healthcare Life Sciences, Pittsbugh, PA) and detected with ECL-2 reagent (Thermo Fisher Scientific). Primary antibodies used in this study were anti-CD9 #13174S (Cell Signaling Technology, Danvers, MA), anti-flotillin-2 #610383 (BD Biosciences, San Jose, CA), anti-CD63 #BDB556019 (BD Biosciences), anti-Ago2 #2897 (Cell Signaling Technology), anti-Dicer #sc-30226 (Santa Cruz Biotechnology), anti-Lupus La #TA500406 (Origene Technologies, Rockville, MD), anti-vinculin #ab129002 (Abcam), anti-nucleophosmin1 #ab10530 (Abcam), anti-nucleolin #ab22758 (Abcam).

Quantitative real-time PCR

Cellular and EV RNAs were extracted with a mirVana miRNA isolation kit (Thermo Fischer Scientific), unless otherwise specified. Taqman miRNA assays for miRNA detection were purchased from Life Technologies. Assay numbers were: hsa-; hsa-miR-486, #001278; hsa-miR-574-3p, #002349; hsa-miR-320a, #002277; hsa-miR-142-3p, #000464; hsa-miR-126, #000451; hsa-mir-145-5p, #002278; hsa-mir-122-5p, #002245; hsa-mir-182-5, 002334. As there are no well-accepted endogenous control transcripts for EVs, the relative quantification was normalized to equal amounts of starting RNA material. RNA was quantified by an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) according to the manufacturer's instructions. RNA (1ng) was used for reverse transcription according to the manufacturer's instructions. Relative quantification was calculated from the expression $2^{-(Ct_{(control)}-Ct_{(experimental)})}$. Taqman qPCR master mix with no amperase UNG was obtained from Life Technologies and quantitative real-time PCR was performed using an ABI-7900 real-time PCR system (Life Technologies).

CRISPR interference

MDA-MB-231 cells expressing dCas9-KRAB, as in (*121*), were generated using lentivirus. A modified version of the transfer plasmid, as in (*121*), UCOE- EF1α-dCas9-BFP-KRAB, was kindly provided by Jonathan Weissman (UCSF). Cells were bulk sorted for blue fluorescent protein (BFP) signal 3 d post transduction and selected cells were expanded by growth for a few generations, and then frozen and stored as parental cells (these cells are referred to as WT throughout the manuscript). Sequences for gRNAs targeting the promoter of the genes of interest were extracted from (*122*). gRNAs were cloned in plasmid pu6-sgRNA EF1Alpha-puro-T2A-BFP (*123*), plasmid #60955 obtained from Addgene, following the cloning protocol as in (*123*). The 3 top gRNAs from the V.2 library (*122*) were chosen per gene of interest. Lentiviruses with the gRNAs targeting the genes of interest were used to transduce the parental cells. Three days post

transduction cells were selected with 2ug/ml puromycin for 3 d. Post puromycin selection, cells were collected for up to 3 generations (~ 72 h) for best levels of protein depletion. More doubling times showed reduced protein depletion.

In vitro packaging reactions

Membrane and cytosol preparation

Fractionation of cells and membranes was done as in (*40*) with some modifications as indicated. MDA-MB-231 cells were harvested at ~80% confluency by adding cold PBS and physically removing the cells by the use of a cell scrapper. Cells were then centrifuged at 1,000xg for 10min at 4°C and cell pellets were frozen at -80°C until use. Cells were thawed and resuspended in 2 volumes of HB buffer (20mM HEPES pH 7,4, 250mM sorbitol) containing protease inhibitor cocktail 1 (1mM 4 aminobenzamidine dihydrochloride, 1 mg/ml antipain dihydrochloride, 1 mg/ml aprotinin, 1 mg/ml leupeptin, 1 mg/ml chymostatin, 1 mM phenymethylsulfonyl fluoride, 50 mM N-tosyl-L-phenylalanine chloromethyl ketone and 1 mg/ml pepstatin). Cells were passed 21-25 times through a 22 gauge needle until >80% of cells were lysed as assessed by microscopy and trypan blue staining. All steps from hereon were done on ice and 4°C, unless otherwise specified. The homogenized cells were centrifuged at 1,500xg for 20 min and the supernatant fraction was subsequently centrifuged at 15,000xg for 15 min to collect donor membranes using a FA-45-30-11 rotor and an Eppendorf 5430 centrifuge (Eppendorf, Hamburg, Germany). The supernatant fraction was centrifuged again at ~150,000xg (49,000 RPM) in a TLA-55 rotor and Optima Max XP ultracentrifuge (Beckman Coulter) to generate the cytosol fraction (5-6 mg/ml). The 15,000xg membrane fraction (pellet) was resuspended in 1 volume of HB buffer and an equal volume of 0.8M LiCl. Donor membranes were then centrifuged again as before and the pellet fraction was resuspended in 1 volume of original starting material HB buffer.

In vitro miRNA packaging reaction

The *in vitro* packaging reaction was performed as described in (*40*). Briefly, miR-122-5p and the variant versions were purchased from Integrated DNA Technologies (IDT, Coralville, IA). Membranes and cytosols were prepared from MDA-MB-231 cells as described above, with the exception of assays using YBX1 KO cells, in which case HEK293T cells WT and YBX1 KO were used as described (*40*). Complete miRNA packaging assays consist of 10ul membranes, 16ul cytosol (the cytosolic concentration was 5.5 mg/ml, unless otherwise specified), 4ul 10X ATP regeneration system (400mM creatine phosphate, 2mg/ml creatine phosphokinase, 10mM AT, 20mM HEPES pH 7.2, 250mM sorbitol, 150mM KOAc, 5mM MgOAc), 8ul 5X incorporation buffer (400 mM KCl, 100 mM CaCl2, 60 mM HEPES-NaOH, pH 7.4, 6 mM, MgOAc), 1ul of 10nM synthetic single stranded RNA and 1ul RNAsin (Promega, Madison, WI). For rescue experiments with purified La, ~0.1ug of total purified La was added per reaction. Reactions were incubated for 20 min at 30°C, then 1ul of RNase If (50,000units/ml) (NEB, Ipswich, MA) and 4ul of 10X NEB buffer 3 (NEB) was added to the reactions and incubated for an extra 20 min at 30°C. Following incubation, RNA was extracted using Direct-Zol (Zymo Research, Irvine, CA) according to manufacturer's instructions and miRNA was quantified using TaqMan miRNA assays as described above. The output was represented as a percentage of protection by comparing the level of miRNA left in

the RNase-treated samples relative to an ice control, which was not RNase-treated. The ice, no RNase control was set to 100%.

Streptavidin pull-down of miR-122 and interacting proteins

The *in vitro* packaging reaction was performed as described above with a modified version of miR-122, 3' biotinylated miR-122 (IDT), which was used to mediate its capture with streptavidin beads. Post-incubation with RNase If, the reactions were heated to 65°C for 15 min to inactivate RNase If and then mixed with 10% TX-100 for a final concentration of 1% and kept on ice for 30 min. Dynabeads Myone Streptavidin T1 (Thermo Fischer Scientific) were washed three times with 1X incorporation buffer and then added to the reactions. The reaction was incubated for 1.5 h at 4°C with constant rotation. Beads were washed with 1X incorporation buffer 5 times and bound proteins were eluted with 1X Laemmli buffer without bromophenol blue. Proteins were electrophoresed in a 4-20% acrylamide Tris-Glycine gradient gel (Life Technologies) for ~3 min. The bulk of proteins were stained with Coomassie and the stained band was excised from the gel using a fresh razor blade. Samples were submitted to the Vincent J. Coates Proteomics/Mass Spectroscopy laboratory at UC Berkeley for in-gel tryptic digestion of proteins followed by liquid chromatography and mass spectrometry analysis according to their standards. The list of detected proteins was then curated for RNA binding proteins, excluding any structural ribosomal protein.

Protein purification

Human La protein was expressed in Sf9 cells as a N-terminal 6X His tagged version. pLJM60-Ssb (*124*), plasmid #38241 obtained from Addgene, served as the La sequence backbone. La was cloned in the pFastBac vector from the Bac-to-Bac Baculovirus expression System (Thermo Fischer Scientific) according to the manufacturer's specifications. The generation of bacmids and the production of baculoviruses was performed as indicated by the manufacturer. For La purification, 1 liter of Sf9 cells was infected with baculovirus. Cells expressing La were harvested 2 d post infection by centrifugation at 1,000xg for 15min at 4°C using a Sorvall RC 6+ centrifuge with a FIBERLite F10-6X500Y rotor. Cellular pellet fractions were stored at -20°C until use. The cellular pellet was thawed and resuspended in 40ml of lysis buffer (20mM HEPES, pH8, 0.1mM EGTA, 500mM NaCl, 5mM DTT, 10mM imidazole, 10mM $MgCl_2$, 250mM sorbitol, 5% glycerol, 1mM PMSF, and protease inhibitor cocktail 1mM 4-aminobenzamidine dihydrochloride, 1 µg/ml antipain dihydrochloride, 1 µg/ml aprotinin, 1 µg/ml leupeptin, 1 µg/ml chymostatin, 1 mM phenymethylsulfonyl fluoride, 50 µM N-tosyl-Lphenylalanine chloromethyl ketone and 1 µg/ml pepstatin)), homogenized and then sonicated. Post-lysis, the material was centrifuged at ~40,000xg (20,000 RPM) using a FIBERlite F211-8X50Y rotor for 20 min at 4°C. The supernatant fraction was centrifuged again at ~42,500xg for 1 h using a Ti70 rotor. In preparation for the sample, 2 ml of Ni-NTA resin beads (Thermo Fischer Scientific) were washed twice with lysis buffer. The supernatant then was mixed with Ni-NTA beads and the suspension incubated with constant rotation at 4°C for 2.5 h. Beads were washed 3 times with washing buffer (recipe for washing buffer was the same as lysis buffer (above) but with 50mM imidazole instead of 10mM imidazole). Proteins were eluted with 4ml of elution buffer (elution buffer similar recipe to lysis buffer, but the imidazole concentration was

250mM). Eluted proteins were applied to a Sephadex G25 column (Thermo Fischer Scientific). Buffer was exchanged to Buffer B (10mM Tris, pH 7,4, 150mM NaCl, 3mM MgCl2, 5% glycerol). Purified fractions were pooled, flash-frozen in liquid nitrogen and stored at -80 C.

Electrophoretic mobility shift assays

In-gel fluorescence was detected in order to assess the free and protein-bound RNA. Fluorescently labeled (5') RNAs (IRD800CWN) were ordered from Integrated DNA Technologies (IDT, Coralville, IA). EMSAS were performed following (*125*), with some modifications. In brief, 1nM of fluorescently labeled RNA was mixed and incubated with increasing amounts of purified La, ranging from 250pM – 2uM. Buffer E was used in this incubation (25mM Tris pH8, 100mM KCl, 1.5mM MgCl2, 0.2mM EGTA, 0.05% Nonidet P-40, 50ug/ml heparin). Reactions were incubated at 30°C for 30 min and then chilled on ice for 10 min and mixed with 6X loading buffer (60mM KCl, 10mM Tris pH 7,6, 50% glycerol, 0.03% (w/v) xylene cyanol). Polyacrylamide gels (6%), acrylamide:bisacrylamide (29:1) (Bio-Rad, Hercules, CA), made with Tris-glycine buffer were prerun for 30min at 200V in the cold room. Samples were resolved in the pre-run gels at 200V for 50 min in a 4°C cold room. Fluorescence was detected by using an Odyssey CLx Imaging System (LI-COR Biosciences, Lincoln, NE). The software of the Odyssey CLx Imaging System was used to obtain quantification of fluorescence. To calculate K_ds, we fit Hill equations with quantified data points. Fraction bound was calculated as a function of exhaustion of free-miRNA.

Immunoprecipitation of La and La-RNA complexes

Approximately 40 million cells were harvested as described in the section "Membrane and Cytosol preparation." Cells were homogenized in 2 volumes of HB buffer and physically disrupted as previously described. Non-lysed cells and nuclei were centrifuged at 1,500xg for 20min. The supernatant fraction was used as a source of cytoplasmic La for immunoprecipitation and was divided into 3 equal parts: input, La-IP, beads only. Dynabead Protein G (Thermo Fischer Scientific) was washed 3 times in polysome lysis buffer (*126*). A 5X polysome lysis buffer was made and mixed with 1,500xg post-nuclear supernatant to a final 1X concentration. Anti-La (Origene Technologies, #TA500406) was added to the post nuclear supernatant to a final concentration of 2ug/500ul and the mixture was incubated with rotation overnight at 4°C. Dynabeads Protein G (Thermo Fischer) beads were added to the mixture and this mixture was incubated for an additional 3 h at 4°C with constant rotation. Beads were washed 5 times with 1X polysome buffer and the content was divided for protein or RNA analysis. Beads for protein analysis were incubated with Laemmli buffer and heated at 95°C for 10 min and beads for RNA analysis were exposed to TRI reagent (Zymo Research) for RNA extraction using Direct-Zol (Zymo Research) as indicated by the manufacturer. Protein and miRNAs were analyzed by immunoblots and TaqMan miRNA qPCR analysis, respectively.

Motif analysis

For motif analysis, we used the MEME Suite 5.0.2 (*127*). MiRNAs detected uniquely in the HD sub-population by TGIRT-seq were pooled together with miRNAs

that were found to be enriched by at least 10-fold (HD/cellular lysate) through TGIRT-seq and Firefly profiling. A total of 49 miRNAs met those requirements and were used for further analysis. A 0 and 1 Markov background model was used to find 3-5 nucleotide motifs. The total mature *Homo sapiens* miRNAs from miRBase were used as background.

RNA extraction and TGIRT-seq library preparation

Cellular and EV RNAs were extracted by using a mirVana miRNA Isolation Kit. Cellular RNA was extracted using a modified version of the mirVana protocol, enriching for <200nt transcripts, in order to reflect EV RNA composition by transcript size. EV RNA was extracted using the standard miRVana protocol. TGIRT-seq libraries were prepared from 2-10ng of starting material.

TGIRT-seq libraries were prepared essentially as described (*94*), with a modification in the starting molecule (see below). Reverse transcription reactions contained purified RNAs, buffer (20 mM Tris-HCl, pH7.5, 450 mM NaCl, 5 mM MgCl2), 5 mM DTT, 100 nM starting molecule (see below) and 1 μM TGIRT-III (Ingex). Reverse transcription by TGIRT-III is initiated by template switching from a starting molecule consisting of a DNA primer (5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATTAN-3') encoding the reverse complement of the Illumina Read2 sequencing primer binding site (R2R) annealed to a complementary RNA oligonucleotide (R2) such that there is a single nucleotide 3' DNA overhang composed of an equimolar mixture of A, G, C and T. The RNA oligonucleotide is blocked at its 3' end with C3Sp (IDT) to inhibit template switching to itself. Reactions were pre-incubated at room temperature for 30 min and then initiated by addition of 1 mM dNTPs. Reactions were then incubated at 60°C for 15 min and terminated by adding 5 N NaOH to a final concentration of 0.25 N and incubated at 95°C for 3 min to degrade RNAs and denature protein. The reactions were then cooled to room temperature and neutralized with 5 N HCl. cDNAs were purified by using a Qiagen MinElute Reaction Cleanup Kit and then ligated at their 3' ends to a modified DNA oligonucleotide (5'Phos-CCTGTTATCCCTAGATCGTCGGACTGTAGAACTCTGAACGTGTAC-3'-C3Sp) encoding the reverse complement of the Illumina Read1 primer binding site (R1R) using Thermostable 5' AppDNA/RNA Ligase (New England Biolabs). Ligated cDNAs were re-purified with MinElute Reaction Cleanup Kit and amplified by PCR for 12 cycles using Phusion DNA polymerase (Thermo Fisher Scientific) with overlapping multiplex and barcode primers that add sequences necessary for Illumina sequencing. PCR reactions were purified using a Select-a-size DNA Clean and Concentrator (Zymo) to remove adapter dimers. Libraries were sequenced on a NextSeq 500 instrument (75-nt, single end reads) at the Genome Sequencing and Analysis Facility at the University of Texas at Austin.

RNA sequencing analysis

Illumina TruSeq adapters and PCR primer sequences were trimmed from the reads with cutadapt 1.16 (*95*) (sequencing quality score cut-off at 20) and reads <15-nt after trimming were discarded. Reads were then mapped using HISAT2 v2.0.2 (*96*) with

default settings to the human genome reference sequence (Ensembl GRCh38 Release 76) combined with additional contigs for 5S and 45S rRNA genes and the E. coli genome sequence (Genebank: NC_000913) (denoted Pass 1). The additional contigs for the 5S and 45S rRNA genes included the 2.2-kb 5S rRNA repeats from the 5S rRNA cluster on chromosome 1 (1q42, GeneBank: X12811) and the 43-kb 45S rRNA repeats that contained 5.8S, 18S and 28S rRNAs from clusters on chromosomes 13,14,15,21, and 22 (GeneBank: U13369). Unmapped reads from Pass 1 were re-mapped to Ensembl GRCh38 Release 76 by Bowtie 2 v2.2.6 (*97*) with local alignment to improve the mapping rate for reads containing post-transcriptionally added 5' or 3' nucleotides (e.g., CCA and poly(U)), short untrimmed adapter sequences, or non-templated nucleotides added to the 3' end of the cDNAs by the TGIRT enzyme (denoted Pass 2). The mapped reads from Passes 1 and 2 were combined using Samtools 1.8 (*98*) and intersected with gene annotations (Ensembl GRCh38 Release 76) supplemented with the RNY5 gene and its 10 pseudogene sequences, which were not annotated in this release, to generate the counts for individual features. Coverage of each feature was calculated by Bedtools (*99*). To avoid mis-mapping reads with embedded sncRNAs, I first intersected reads with sncRNA annotations and the remaining reads were then intersected with the annotations for protein-coding genes, lincRNAs, antisense, and other lncRNAs. To further improve the mapping rate for tRNAs and rRNAs, I combined reads mapped to tRNAs or rRNAs in the initial alignments and re-mapped them to tRNA reference sequences (Genomic tRNA Database, and UCSC genome browser website) or rRNA reference sequences (GeneBank: X12811 and U13369) using Bowtie 2 local alignment. Because similar or identical tRNAs with the same anticodon may be multiply mapped to different tRNA loci by Bowtie 2, mapped tRNA reads were combined according to their anticodon (N = 48) prior to calculating the tRNA distributions.

For miRNA analysis, sequences were mapped to miRbase using miRdeep2. Reads were normalized by dividing the numbers or reads per miRNA to the total number of miRNA reads in each sample and then this value was multiplied by one million (reads per million miRNA mapped reads – RPM).

Results

MiR-122 sorting is recapitulated in a cell-free reaction

Due to the roles of circulating miR-122 in breast cancer biology (prognostic biomarker for metastasis (*109*) and promoter factor of invasiveness (*32*)), I selected miR-122 for further biochemical analysis. To explore the mechanism(s) by which miR-122 is selectively sorted into the MDA-MB-231 HD EV sub-population, I first sought to recapitulate miR-122 packaging *in vitro*. An assay for the cell-free packaging of a miRNA into putative exosomal vesicles was previously developed in our laboratory (*40*). The reaction includes sedimented membranes and clarified cytosol, obtained from mechanically ruptured HEK293T cells, along with synthetic miRNA and ATP. After an incubation of 20 min at 30°C, RNA incorporated into vesicles was monitored by protection against degradation by exogenous RNase. Protected RNA was then extracted and quantified by RT-qPCR. The outcome of the reaction is represented as the percentage of initial input miRNA that has become protected over the incubation period (Figure 3-1A). Using a modified version of our previously published reaction, I found that

maximum protection of miR-122 in lysates of MDA-MB-231 cells required membranes, cytosol and incubation at a physiological temperature (Figure 3-1B), as was previously observed for miR-223 in lysates of HEK293T cells (*40*). Therefore, miR-122 packaging into vesicles in lysates of MDA-MB-231 cells may be recapitulated *in vitro.*

Identifying *in vitro* miR-122 RNA-binding protein partners

Having detected miR-122 packaging *in vitro*, we next sought to utilize this approach to study the mechanism of miR-122 sorting into MDA-MB-231 EVs. I first tested if miR-122 packaging was YBX1-dependent. YBX1 is a RNA-binding protein (RBP) previously found to be required for miR-223 packaging into vesicles *in vitro* and into EVs secreted by HEK293T cells (*40*). Using membranes and cytosol from WT and *ybx1* null HEK293T cells, I repeated the demonstration that miR-223 packaging was YBX1-dependent, but found that miR-122 was packaged nearly normally in lysates devoid of YBX1 protein (Figure 3-2A).

Several RBPs have been implicated in miRNA sorting into EVs from different cell types (*40, 42, 43, 128*). As such, in order to study the RBP(s) that might mediate miR-122 packaging in MDA-MB-231 cells, I performed an *in vitro* packaging reaction employing a 3'-biotinylated form of miR-122 to allow the capture of the miRNA and any bound proteins. Briefly, following *in vitro* packaging, reactions were treated with RNase, the RNase activity was quenched and the membranes solubilized with Triton X-100. Once the luminal content was released, biotinylated miR-122, along with its protein interactors, was captured with streptavidin beads. Proteins were eluted with Laemmli buffer, extracted from a SDS-polyacrylamide gel and the eluted fraction used for mass spectrometry. The proteins detected by mass spectrometry were curated for RBPs, except that any ribosomal proteins were excluded (Figure 3-2B). I decided to focus on the top three candidates, nucleolin (NCL), Lupus La (La) and nucleophosmin (NPM1). These three RBPs have been reported to be present in crude high-speed pellet preparations from conditioned media from different carcinoma cell lines (*129, 130*), including from our model cell line MDA-MD-231 (*131*).

I next examined the requirements for co-packaging of the top three RBP candidates and miR-122 in the cell-free reaction. An immunoblot was used to test the relative amount of La, nucleophosmin and nucleolin co-isolated with biotinylated miR-122 in conditions that sustain or fail to result in miR-122 packaging. I observed maximal La, nucleophosmin and nucleolin recovery in incubations conducted with cytosol and membranes at 30°C but much less so at 4°C (Figure 3-2C). Thus, the sorting of the top RBP candidates bound to miR-122 requires the presence of both, membranes and cytosol, and a physiological temperature of incubation.

Notably, Ago2 was not detected bound to miR-122 in our mass spectrometry results. This was not simply an artifact of our *in vitro* system, as Ago2 was also undetectable in immunoblots of the buoyant density fractionated EV membranes (Figure 2-1C). Both, Ago2 and Dicer were present in the high-speed pellet, but not as buoyant species, suggesting they are associated with co-purifying RNP complexes that are not vesicle-associated. This finding is in accordance with other published data (*23, 40, 41*)

where Ago2 is detected in the high-speed pellet but absent in the vesicle sample after more stringent purification methods are used.

These results suggest a potential role of La, nucleophosmin and nucleolin on miR-122 packaging *in vitro*.

MiR-122 *in vitro* packaging is La-dependent

To test the relevance of the three RBPs in miR-122 packaging into EVs, I used CRISRPR interference (CRISPRi) (*121, 122*) to systematically knock down each protein in MDA-MB-231 cells. CRISPRi promotes gene silencing by repressing transcription of the target gene. Importantly, unlike siRNA or shRNA, CRISPRi silences genes independently of the RISC machinery. Because the RISC machinery binds to miRNAs and is responsible for miRNA-mediated gene silencing (*28*), I avoided any artificial overload of the RISC machinery that might result in unpredictable effects on the miRNA sorting in our system.

Using CRISPRi, I prepared cytosols from MDA-MB-231 cells depleted of nucleophosmin and La that were then used in the *in vitro* packaging reactions. Both nucleophosmin and La were efficiently knocked down using this system (Figure 3-3A). However, nucleolin knock-down resulted in apparent cellular arrest with subsequent cell death. Thus, nucleolin functional analysis was not possible by knock-down. Nucleophosmin-depleted cytosol produced a mild packaging phenotype only at reduced cytosol concentrations (3 mg/ml) (Figure 3-3B). In contrast, La-depleted cytosol showed severely reduced miR-122 packaging efficiency at cytosolic concentrations of 3 and 6 mg/ml in the cell-free reactions (Figure 3-3B, C). Thus, these experiments suggested a role for La in the packaging of miR-122 into vesicles formed *in vitro.*

La is an abundant RNA-binding protein that shuttles between the nucleus and cytoplasm. Within the nucleus, La helps stabilize RNA polymerase III (PolIII) transcripts by binding to poly-uridine tails in their 3' termini (*132-134*). In the cytoplasm, La has been suggested to bind and regulate the translation of 5' terminal oligopyrimidine (TOP) mRNAs (*135-137*). Moreover, in the context of cancer, La has been shown to shuttle from the nucleus to the cytoplasm at higher efficiencies in invasive cells undergoing epithelial to mesenchymal transition in order to control the translation of mRNAs bearing internal ribosome entry sites (IRES) (*138, 139*). La has also been indirectly implicated in the exosomal secretion of a small Epstein-Barr viral RNA, EBV-EBER1. EBV-EBER1 interaction with La in the cytoplasm masks the viral RNA from recognition by cytoplasmic surveillance machinery and facilitates its later EV-mediated secretion (*140*).

In order to establish whether La itself, rather than a La-associated protein, is necessary for miR-122 incorporation into EVs *in vitro* (Figure 3-3C), recombinant His-tagged human La was expressed and purified from insect Sf9 cells (Figure 3-3D). La-depleted cytosol was reduced ~4-fold in the packaging of miR-122 but was restored to normal on addition of purified, recombinant La to the combination of membranes and La-depleted cytosol (Figure 3-3E). However, addition of La alone with or without membranes and no cytosol resulted in only background protection of miR-122 (Figure 3-

3E), suggesting that additional cytosolic proteins are necessary for *in vitro* vesicle biogenesis and/or miRNA sorting process. The level of exogenously added La protein in the *in vitro* reaction was adjusted to approximate the level found in an aliquot of the normal cytosol (Figure 3-3F). I thus demonstrated that the La protein is required for miR-122 packaging *in vitro* in MDA-MB-231 cells.

MiR-122 packaging into EVs *in vivo* is La-dependent

The *in vitro* data suggested that La was at least one of the RBPs responsible for miR-122 packaging. However, in order to confirm this conclusion, I first examined the effects of depletion of the three RBP candidates *in vivo*. Previous work has demonstrated that selected miRNAs accumulate in the cytoplasm in cells deficient in RBPs required for RNA sorting into EVs (*40, 42, 43*). I assessed the level of intracellular miR-122 and of another highly enriched miRNA (miR-142) upon CRISPRi depletion of each RBP candidate. Greater accumulation of miR-122 was seen upon La depletion (2.7-fold change La CRISPRi/WT) compared to nucleophosmin depletion (1.7-fold change NPM1 CRISPRi/WT) (Figure 3-4A). CRISPRi of nucleolin resulted in less cellular miR-122, which could be an indirect result of the more toxic effects of this depletion. As a control for selectivity of miRNA sorting and export, I assessed the intracellular levels of non-selectively sorted miRNAs miR-574 and miR-320a (Figure 2-4H). In contrast to the selectively sorted miRNAs, I did not observe intracellular accumulation of miR-574 and miR-320a under any of the knock-down conditions (Figure 3-4). At least in relation to La and nucleophosmin, the depletion results in cells confirmed our previous observations on the role of these proteins in our cell-free reaction.

I next examined the relative content of miR-122 and of a non-selectively sorted miRNA, miR-320a, in EVs in relation to their accumulation in cells depleted of La. WT and La-depleted cells had approximately the same number of purified EVs as quantified using a Nanosight particle tracking instrument (not shown observations). A significant reduction in secretion of miR-122 with no change in miR-320a was accompanied by intracellular accumulation of miR-122 in the La-depleted background (Figure 3-4B). Thus, La plays a role in the selective sorting of miR-122 into EVs, without affecting EV biogenesis.

To test for the presence of La in EVs isolated from cultured MDA-MB-231cells, I first purified vesicles by buoyant density flotation. The high-speed pellet was subjected to flotation in a step sucrose gradient. Buoyant membranes were permeabilized with non-ionic detergent and soluble proteins were precipitated/concentrated with acetone in order to facilitate detection by immunoblot (Figure 3-4C). I successfully detected the presence of La in the buoyant sample. Its detection was improved upon acetone precipitation/concentration (Figure 3-4D). Moreover, when I analyzed the La distribution across the iodixanol linear gradient (as in Figure 2-1A), I observed La co-fractionating with the HD sub-population (Figure 3-4E). I previously showed that the majority of secreted miR-122 co-fractionates with the HD sub-population (Figure 2-5E). Those two observations are in agreement for the role of La in the secretion of miR-122. Interestingly, nucleolin, one of the other RBP candidates that pulled-down with biotinylated miR-122 in the *in vitro* packaging reaction (Figure 3-2B) was also detected

in the HD fraction (Figure 3-4E). Nucleolin may also be involved in RNA secretion into HD transcripts.

In order to confirm that La was a soluble protein residing inside EVs, and not simply vesicle-associated, I performed proteinase K protection assays on crude high-speed pellet material. The high-speed pellet fraction from conditioned medium was exposed to proteinase K in the presence or absence of detergent and the exposure of La was assessed by immunoblot. The La protein was resistant to proteinase K digestion and rendered sensitive upon membrane permeabilization with non-ionic detergent (Figure 3-4F). Flotillin-2, a membrane protein associated with the inner leaflet of the EV, served as a positive control and was degraded only in the presence of detergent. In contrast, Dicer, previously shown to be present in the high-speed pellet but not as an EV-encapsulated RNP (Figure 2-1C), was susceptible to degradation independent of the addition of non-ionic detergent (Figure 3-4F). We conclude that the La protein is packaged into vesicles in our cell-free reaction as well as into EVs secreted by cells and is required for packaging of miR-122 into vesicles *in vitro* and *in vivo*.

La directly interacts with miR-122 *in vivo* and *in vitro*

I then evaluated the interaction of La with two miRNAs by co-immunoprecipitation from mechanically ruptured MDA-MB-231 cells. A post-nuclear supernatant fraction of cells was mixed with La antiserum and the IP sample was split for protein and RNA analysis by immunoblot and RT-qPCR, respectively (Figure 3-5A). Approximately 15% of the La protein and miR-122 was precipitated (Figure 3-5B, C). MiR-182, which is present but not enriched in LD and HD EVs, was used as a negative control and did not co-immunoprecipitate with La (Figure 3-5C).

In order to explore the interaction of La and miR-122 directly, I used purified, recombinant La incubated with miR-122 and evaluated the binding affinity by electrophoretic mobility shift assays (EMSAs). Purified La was titrated, mixed with 5' fluorescently-labeled miR-122 and incubated at 30°C. After incubation, complexes were separated by electrophoresis and detected by in-gel fluorescence (Figure 3-5D). As negative and positive controls for La interaction, I used 22 nt RNAs comprised of alternating purines or poly-uridine, respectively (Figure 3-5D). The measured K_d for La:miR-122 was 4.8 nM (Figure 3-5E). Notably, this measurement indicates that the affinity of miR-122 for La is greater than the affinity previously reported for Ago2 and miRNAs, with K_ds from 10-80 nM (*141*). We therefore conclude that the La:miR-122 complex displays high affinity and specific interaction *in vitro* and *in vivo.*

Finding the miR-122 motif responsible for its binding and packaging

I then explored the miR-122 motif responsible for its association with La and packaging into EVs in the cell-free reaction. La has been shown to bind to UUU sequences at the 3' end of PolIII transcripts (*132*) and miR-122 contains a UUUG at that terminus (Figure 3-6A). I designed a variant of miR-122 where A residues replaced the 3'U bases and found that this species is poorly packaged into vesicles in the cell-free reaction (3% efficient incorporation vs. 15% for WT) (Figure 3-6B) and displays a greatly reduced (~100-fold) affinity for La in the EMSA analysis (WT K_d 4.8 nM vs

3'mut K_d 538 nM) (Figure 3C, D). We conclude that the UUU sequence at the 3' end of miR-122 is necessary for a high affinity interaction with La, which may correlate to the requirement for these residues in the packaging of miR-122 into vesicles in the cell-free reaction.

To identify possible motifs beyond the literature-based UUU candidate, I also performed a multiple expectation maximization (EM) for motif elicitation (MEME) analysis for 49 HD EV miRNAs. I considered miRNAs that were at least 10-fold enriched, as detected by profiling and sequencing, as well as those unique miRNAs in the HD sub-population that were detected by sequencing. This approach identified "UGGA" as an EV-enriched miRNA motif. This motif was present in 13 (26%) of the 49 analyzed miRNAs, including the 5' sequence of miR-122 (Figure 3-6A), and it was by far the most overrepresented motif detected by this analysis. The UGGA sequence appeared to be EV-specific, as it did not emerge when similar analyses were performed with exclusively cellular miRNAs (miRNAs detected in cellular lysates but absent in the EV samples). I then manually assessed the pool of exclusively cellular miRNAs for this motif where it appeared in 5 out of 97 analyzed miRNAs (5%). A variant of this 5' sequence motif impaired miR-122 incorporation in the *in vitro* packaging reaction (Figure 3-6B) and decreased the affinity ~10-fold for La compared to the WT sequence (Figure 3-6C, D). Thus, we conclude that a bipartite, or possibly even more complex motif directs recognition of miR-122 by La and consequently results in its selective sorting into EVs and secretion from MDA-MB-231 cells.

The La protein is required for miRNA and TOP mRNA sorting

I demonstrated that the La protein is required for miR-122 secretion in vitro and in vivo. I then asked whether La plays a broader role in the selective sorting of HD miRNAs. I tested the secretion of validated HD miRNAs (Figure 2-4G) that proved to reside inside vesicles (Figure 2-5A) and the secretion of non-selectively sorted miRNAs (Figure 2-4H) under La depletion vs. WT conditions. All the selectively sorted miRNAs tested showed some dependence on La for their secretion, as seen by at least a 50% reduction of their secretion into EVs in comparison to WT levels (Figure 3-7A). By contrast, all of the non-selectively sorted miRNAs tested showed little to no change under La depletion (Figure 3-7A). We then performed TGIRT-seq in EVs derived from WT or La depleted cells. Analysis of miRNAs in WT EVs ranked by abundance showed similar results as the RT-qPCR analysis, with multiple miRNAs having an impaired secretion upon La depletion (Figure 3-7B). The miRNAs having an impaired secretion as detected by TGIRT-seq include miR-142, miR-122, miR-126 and miR-486, confirming our RT-qPCR results.

La has been suggested to bind and regulate the translation of 5' terminal oligopyrimidine (TOP) mRNAs in the cytoplasm (*135-137*). Previously we found that 5'TOP mRNAs are enriched in EVs derived from HEK293T cells in comparison to their relative cellular levels (*26*). Therefore I analyzed the TGIRT-seq data to evaluate the status of the secreted 5' TOP mRNAs under La depletion conditions. An overall reduction of 5' TOP mRNAs in EVs derived from the La depletion background was observed (Figure 3-7C) without an apparent change in 5' TOP mRNA abundance in cells

(Figure 3-7D). Secretion of tRNAs, the most prominent EV transcripts (Figure 2-3B), remained unchanged under La depletion conditions (Figure 3-7C), despite their slight drop on abundance in cells (Figure 3-7D). The diminished presence of tRNAs in La depleted cellular lysates is in agreement with previous studies implicating the La protein as an important player on tRNA maturation and stabilization (*134, 142*).

All of these together show a role for La in the secretion of selectively sorted miRNAs and 5' TOP mRNAs into EVs. The sorting of other major EV transcripts, such as tRNAs may be a La-independent process.

Discussion

Here I document the requirement of the La protein for the EV-mediated secretion of miR-122. By using genetic and biochemical tools, I find that the La protein binds with high specificity and affinity to miR-122 and the La-miR-122 complex is secreted in HD vesicles (Figure 3-8). Moreover, I show that the La protein is also required for the secretion of selectively sorted miRNAs and 5' TOP mRNAs.

The La protein is an abundant RBP that is upregulated in some cancers (*143-146*). Its presence in the cytoplasm has been linked to the translational regulation of mRNAs bearing IRES (*138, 139*), as well as 5'TOP mRNAs (*135-137*). Interestingly, it was previously reported that 5'TOP mRNAs are enriched in EVs derived from HEK293T cells in comparison to their relative cellular levels (*26*). The secretion of 5'TOP mRNAs in HEK293T is not affected by the depletion of the RNA binding protein YBX1. YBX1 did affect the secretion of other major EV transcripts such as tRNAs and YRNAs. Here I show the dependence of La for the secretion of 5' TOP mRNAs (Figure 3-7C) in MDA-MB-231 cells.

Little is known about how cytoplasmic RNA-binding proteins, such as La, along with bound RNA molecules are targeted and recognized for secretion. The phosphorylation status of La has been suggested to be important for its intracellular localization (*135*). In addition, sumoylation of La has been shown to be important for La-dependent mRNA transport in neurons (*147*) and sumoylation of another RNA-binding protein in EVs (hnRNPA2B1) has been shown to be important for export (*42*). Post-translational modifications such as phosphorylation, ubiquitylation and sumoylation may serve to direct RNPs to the ESCRT machinery and thus into membrane buds that are internalized into MVBs.

The La protein binds to and stabilizes poly-uridine stretches of PolIII transcripts that mature in the nucleus (*134*). Likewise in the cytoplasm, La may serve a similar function in the stabilization of polyuridine tracks in 5'TOP mRNAs (*135*). Here I show that La directly binds to, and promotes the membrane enclosure of miR-122 in a cell-free reaction (Figure 3-3E). I obtained evidence that a 3' poly-uridine sequence of miR-122 is required for this interaction (Figure 3-6C, D). This role of La may extend to miRNAs secreted by other cell lines. B cell-derived EVs are enriched for 3' uridylylated miRNA isoforms in comparison to their intracellular counterparts (*148*), and La has been detected

by mass spectrometry in the high-speed pellet fraction from B cell conditioned medium (*149*).

In addition to the known poly (U) binding site, sequence motif analysis of the most highly enriched miRNAs in the HD vesicles isolated from MDA-MB-231 conditioned medium identified a second, previously unreported motif required for La:miR-122 high affinity interaction and sorting into EVs in our cell-free reaction. A 5' UGGA sequence resembles the motif shown to be required for miRNA sorting into hepatocyte EVs, dependent on the SYNCRIP protein (*43*). However, I did not detect SYNCRIP in our mass spectrometry analysis for RBP partners responsible for miR-122 sorting *in vitro* (Figure 3-2B). Thus, it is possible hepatocytes and breast cancer epithelial cells differ in their mechanisms of miRNA sorting. The presence of both motifs, UUU and UGGA, in miR-122 may allow for an exceptionally high affinity interaction with La. The measured K_d for the La:miR-122 complex is 4.8 nM (Figure 3-5E), indicating tighter binding than that reported for Ago2 to miRNAs using a similar assay (*141*). This high affinity interaction may explain how selected miRNAs are sequestered into exosomes away from RISC.

As we have shown with La, other RBPs (YBX1 (*40*), hnRNPA2B1 (*42*) and SYNCRIP (*43*)) are involved in the selective sorting and secretion of miRNAs in different cell lines. Conceivably, this process could be organized and controlled in a cell-specific manner by the assembly of RNA granules including other more abundant RNAs such as tRNA, Y-RNA and vault RNA (*26*) (Shurtleff et al., 2017). Studying the role of RNA granules in the context of RBP and RNA sorting into EVs will be instructive in the future.

The physiological reason behind the selective miRNA and 5' TOP mRNA sorting mediated by the La protein is unknown. One alternative is that metastatic cells sort these transcripts into vesicles for intercellular communication to promote their invasiveness. Delivered miRNAs could change gene expression in recipient cells to promote metastasis, e.g. by downregulating glucose or metabolite transport machinery in recipient cells, so the metastatic EV producing cells have a better availability of nutrients. Moreover in the context of a tumor, it is possible that cancer cells package and deliver functional 5' TOP mRNAs to neighboring cells. This would help tumor development by promoting the growth and metabolic state of the recipient cells. Alternatively the secretion of miRNAs and 5' TOP mRNAs could serve cancer cells as a way to dispose unwanted material, such as tumor-suppressor miRNAs. Furthermore in the context of cancer cells coping with stressors such as starvation, disposing 5' TOP mRNAs could help them reaching a homeostatic state.

Figures

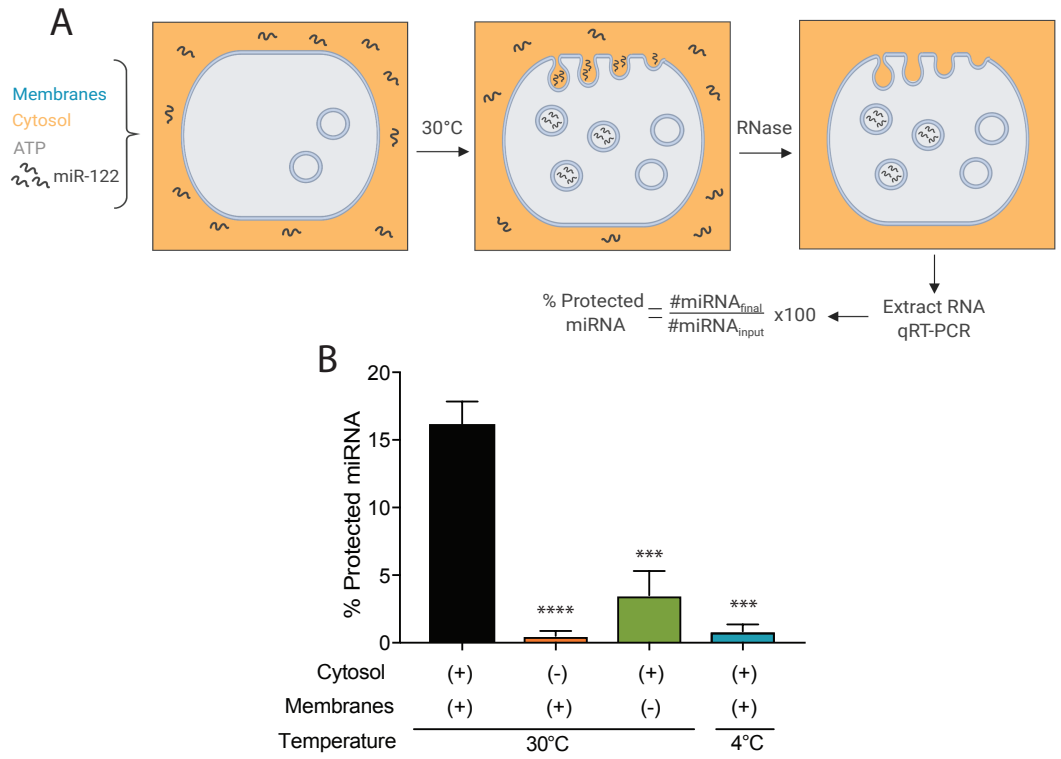
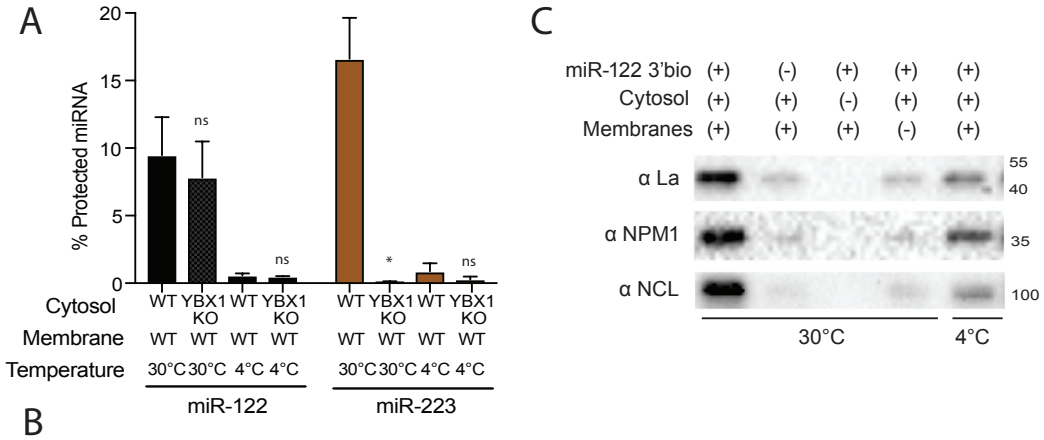


Figure 3-1: miR-122 packaging is recapitulated in a cell-free reaction.

(A) Schematic depicting the *in vitro* packaging reaction. Image reproduced from the original manuscript developing the cell-free reconstitution assay (*40*). (B) Quantification of the *in vitro* packaging reaction of miR-122. Reactions with or without membranes (15,000xg pellet), cytosol (150,000xg supernatant) prepared from MDA-MB-231 cells and incubated at 30°C or 4°C are shown. Data plotted are from 3 independent experiments, each independent experiment with triplicate qPCR reactions; error bars represent standard deviation from independent samples. Statistical significance was performed using Student's t-test (***p<0.001, ****p<0.0001).



B

Protein	# of unique peptides	Coverage
Nucleolin	13	23.5%
Lupus La	7	25.7%
Nucleophosmin1	7	35.0%
Poly (rC)-binding protein	4	14.6%
Small nuclear ribonucleoprotein Sm D1	3	25.2%

Figure 3-2: Identifying *in vitro* miR-122 RNA-binding protein partners.

(A) Quantification of the in vitro packaging of miR-122 and miR-223. Cytosols from WT or YBX1 KO HEK293T cells were used. Incubations at 30°C or 4°C are shown. Data plotted are from 2 independent experiments, each independent experiment with triplicate RT-qPCR reactions; error bars represent standard deviation from independent samples. (B) List of RNA-binding proteins pulled down with biotinylated miR-122 after a modified in vitro packaging reaction was performed (see materials and methods). (C) Immunoblots for La, NPM1 and NCL following miR-122 *in vitro* packaging performed as for Figure 3-1B according to the conditions indicated. Statistical significance was performed using Student's t-test (*$p<0.05$, ns = not significant).

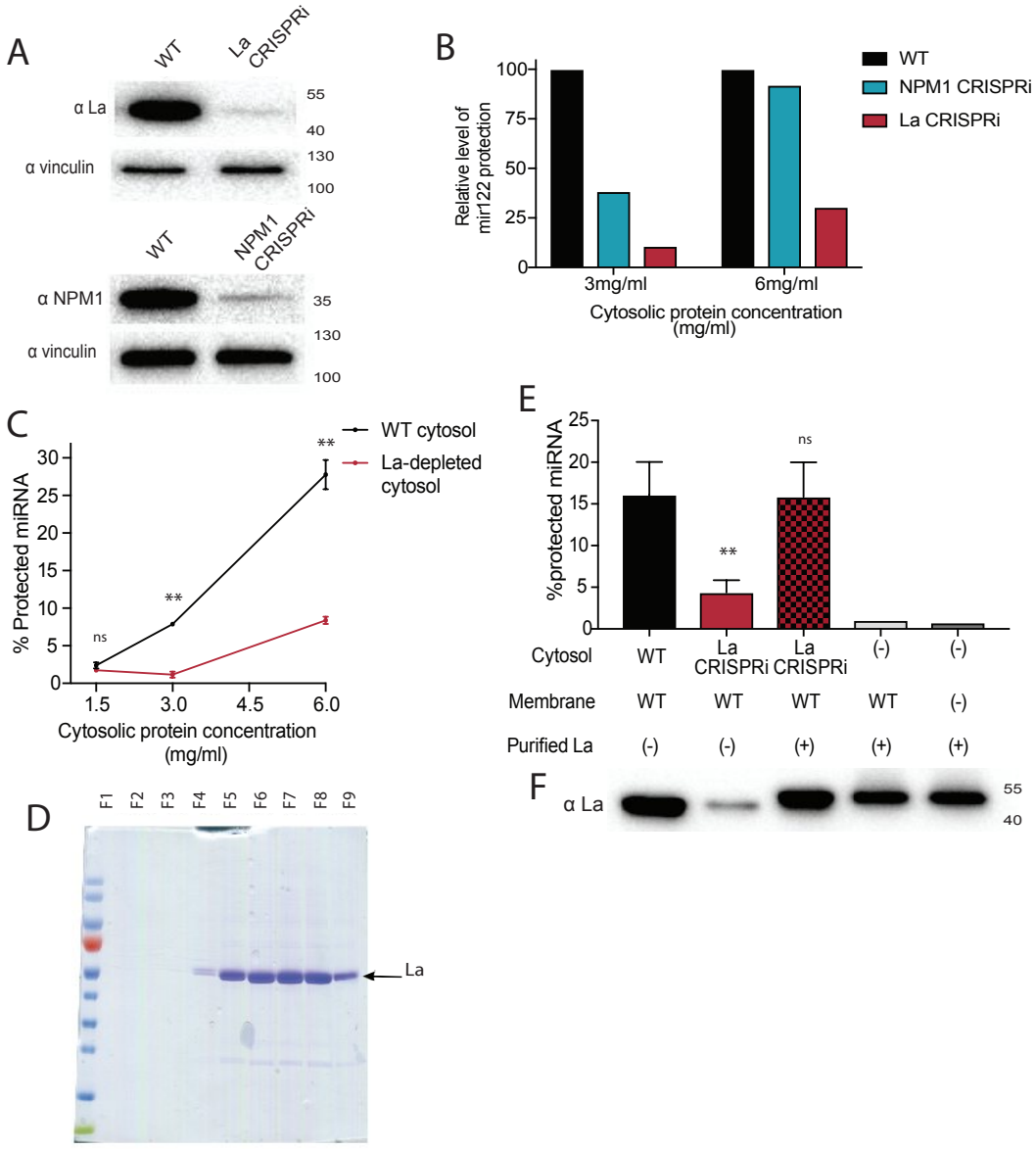


Figure 3-3: miR-122 *in vitro* packaging is La-dependent.

(A) Analysis of RBP depletion by immunoblots upon CRISPRi depletion. Immunoblots for La, NPM1 and vinculin are shown. (B) Relative quantification of miR-122 *in vitro* packaging. Cytosols from WT, La or NPM1-depleted backgrounds were tested in the reactions (cytosols from A were used). Two cytosolic protein concentrations were tested: 3 and 6mg/ml. (C) Quantification of miR-122 *in vitro* packaging. WT and La-depleted cytosols were titrated from 1.5 to 6mg/ml. Data plotted are from 2 independent experiments, each independent experiment with triplicate qPCR reactions; error bars represent standard deviation from independent samples. (D) Purification of heterologously expressed La. Coomassie staining of purified La fractions. Fractions shown were collected post gel filtration on Sephadex G25 columns. (E) Heterologously expressed La rescues miR-122 *in vitro* packaging. Quantification of miR-122 *in vitro* packaging is shown. Cytosols from WT or La-depleted cells with or without

complementation using purified La were used. Reactions without cytosol, with or without membranes and with added purified La are also shown. Data plotted are from 3 independent experiments, each independent experiment with triplicate qPCR reactions; error bars represent standard deviation from independent samples. (F) Immunoblots showing the levels of endogenous or heterologously expressed added La used in the *in vitro* reactions as in (E) are shown.

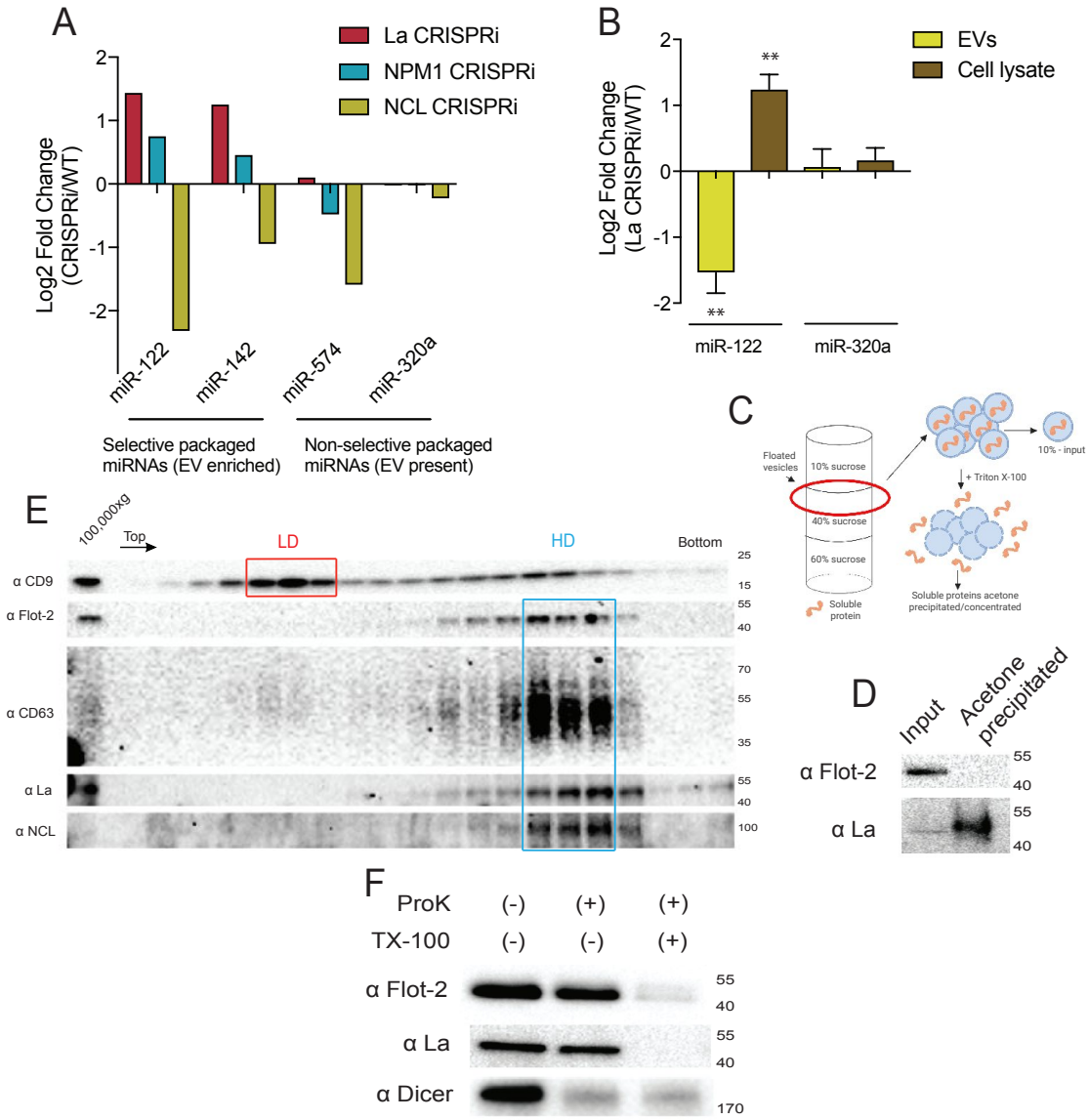


Figure 3-4: miR-122 packaging into EVs *in vivo* is La-dependent.

(A) Log2 fold change for intracellular levels of miRNAs of interest post-depletion of the shown RBPs by CRISPRi. MiR-122 and miR-142 were used as representatives of selectively packaged miRNAs. MiR-574 and miR-320a were used as representatives of non-selectively packaged miRNAs. Data quantified by RT-qPCR, normalized per ng of total RNA. (B) Log2 fold change of intracellular and secreted levels of miRNAs of interest quantified by RT-qPCR. EVs were purified from WT and La-depleted cells. Cellular lysates were isolated at the moment of EV collection. MiR-320a was used as a control for non-selectively packaged miRNAs. Data plotted are from 3 independent experiments (biological replicates); each independent experiment with triplicate qPCR reactions, error bars represent standard deviation from independent samples. (C) Schematic representation of the flotation assay and acetone precipitation as in D. The high-speed pellet of conditioned medium was floated in a step sucrose gradient (input) and the soluble content was released by the addition of TX-100 and then concentrated by

acetone precipitation (acetone precipitated). (D) EV samples post-flotation and acetone precipitation/concentration were tested. Immunoblots for flotillin-2 and La are shown. Acetone precipitation captures mostly soluble proteins (*150*), thus flotillin-2 is only detectable in the floated sample, prior to acetone precipitation, whereas La detection improves post precipitation/concentration. (E) The La protein co-fractionates with the HD sub-population. Immunoblots across the Optiprep linear gradient for classical EV markers as well as for La and nucleolin are shown. (F) Proteinase K protection assays in high-speed pellet fractions. Samples were treated with or without proteinase K and or Triton X-100. Immunoblots for flotillin-2, La and Dicer are shown. Statistical significance was performed using Student's t-test (**$p<0.01$).

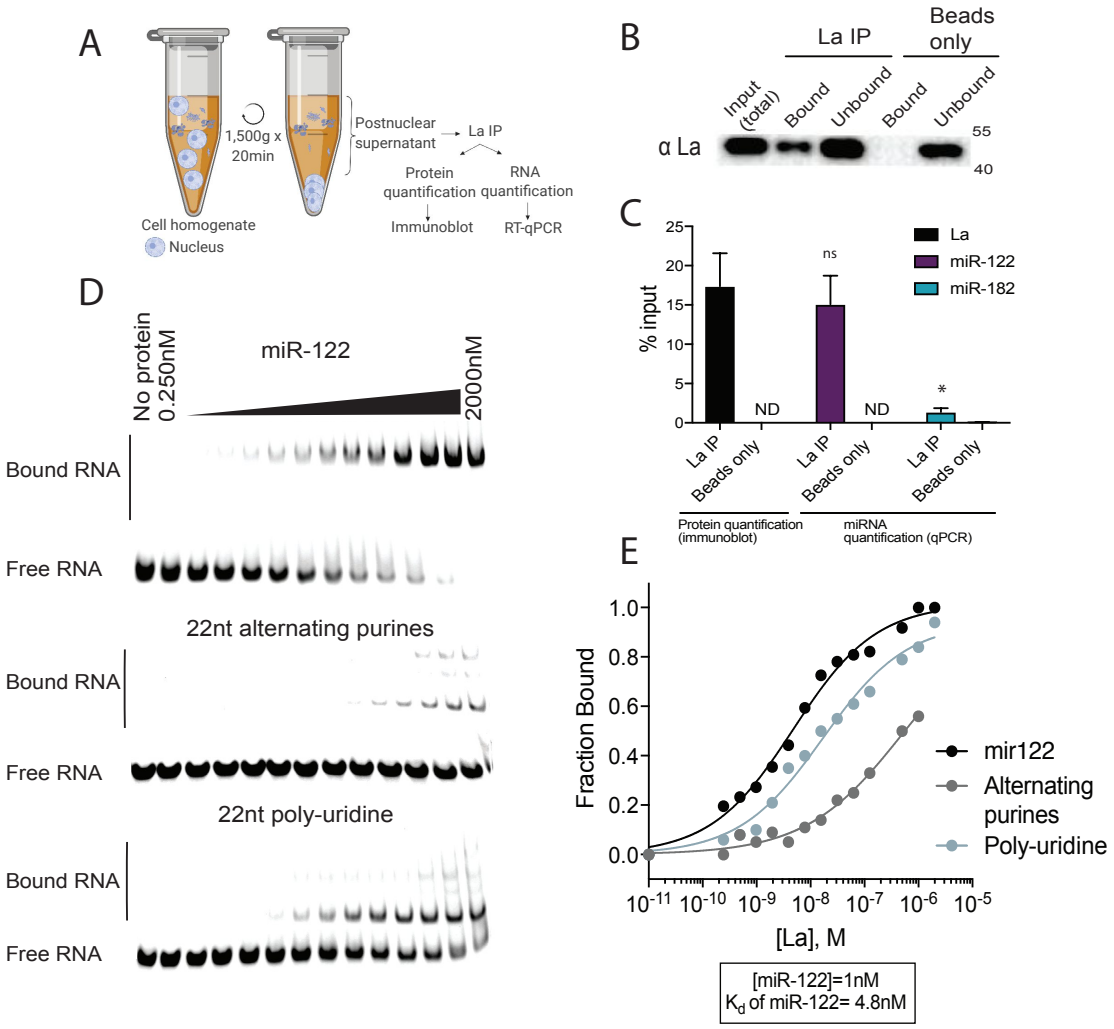


Figure 3-5: La interacts directly with miR-122 *in vivo* and *in vitro*.

(A) Schematic representing immunoprecipitation of endogenous La shown in B and C. (B) Immunoblot for La post-La IP. (C) Quantification of the La immunoblot post La IP. The relative levels, as percentage of input, of miR-122 and miR-182 co-IP with La are also shown. MiR-182, a non-selectively sorted miRNA, served as a negative control for La binding. Data plotted are from 2 independent experiments (biological replicates), for qPCR data each independent experiment with triplicate qPCR reactions; error bars represent standard deviation from independent samples. (D) EMSA assays using 5' fluorescently labeled miR-122, a 22nt RNA consisting of alternating purines (negative control for binding) and a 22nt poly-uridine oligonucleotide (positive control for binding). Purified La was titrated from 250pM to 2uM. In gel fluorescence was detected. (E) Quantification of D showing the calculated K_d. Fraction bound was quantified as a function of exhaustion of free miRNA. Statistical significance was performed using Student's t-test (*$p<0.05$, ns = not significant).

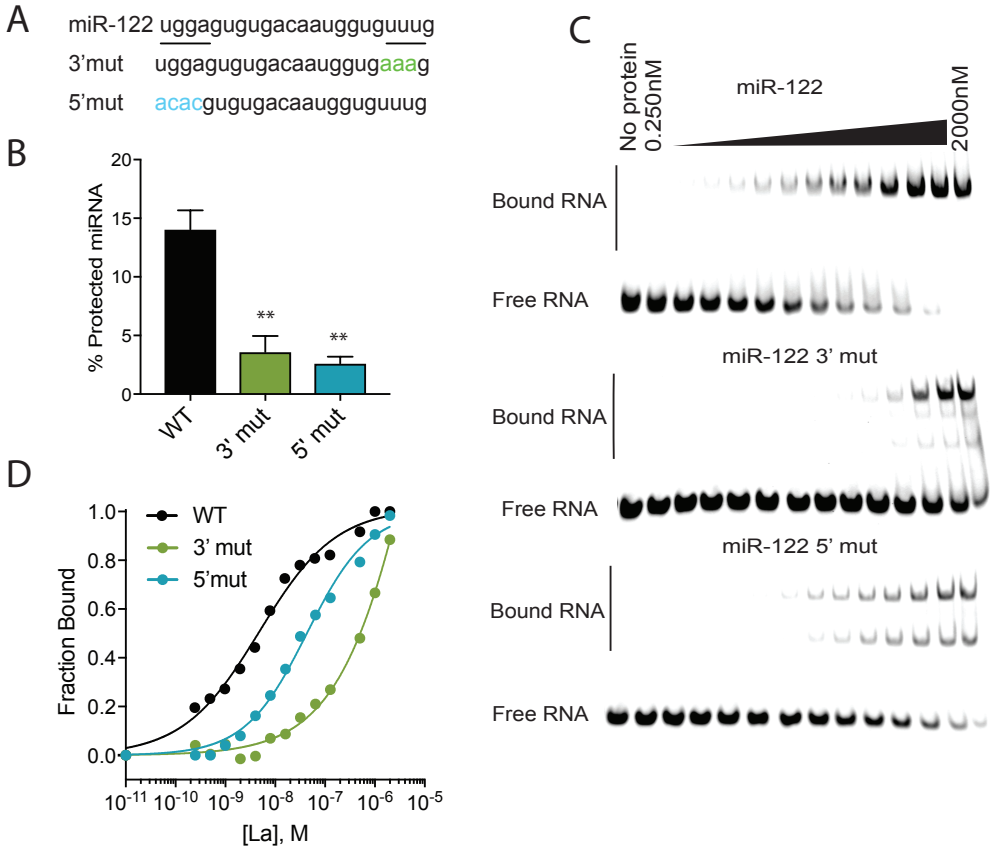


Figure 3-6: A bipartite motif in miR-122 is required for its packaging and interaction with La *in vitro*.

(A) The sequences of miR-122 WT and mutated versions are shown. (B) Quantification of *in vitro* packaging of miR-122 WT and mutated versions. Data plotted are from 3 independent experiments, each with triplicate qPCR reactions; error bars represent standard deviation from independent samples. (C) EMSAs for miR-122 WT and the mutated versions. The distinct miR-122 versions were incubated with increasing amounts of La. (D) Binding affinity curves (as calculated from EMSAs as in B) for miR-122 and the mutated versions are shown. Fraction bound was quantified as a function of exhaustion of free miRNA. Statistical significance was performed using Student's t-test (**$p<0.01$).

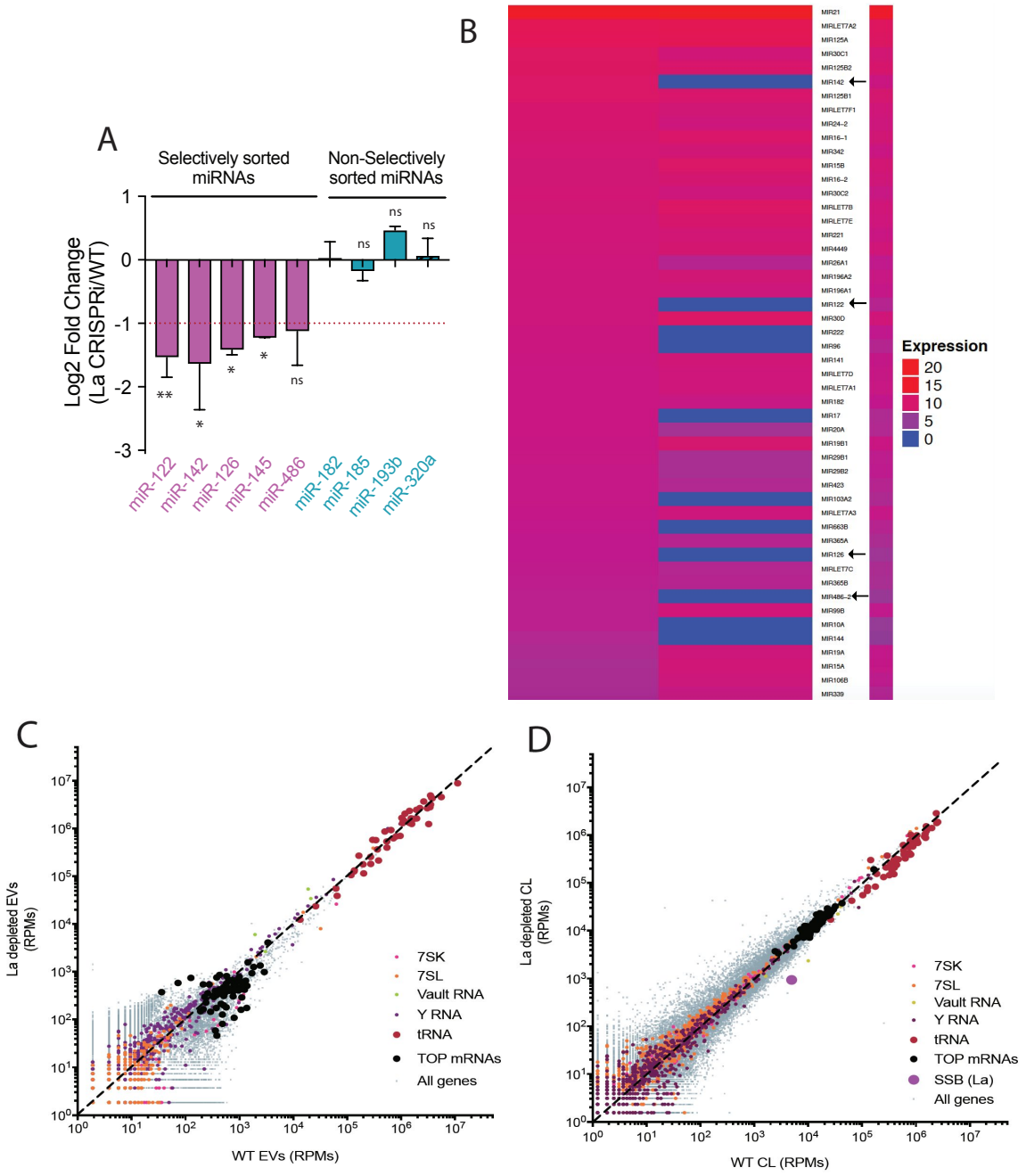


Figure 3-7: The La protein is required for selective miRNA and 5' TOP mRNA sorting.

(A) Log2 fold change of secreted levels of miRNAs of interest quantified by RT-qPCR. EVs were purified from WT and La-depleted cells. MiRNAs in pink represent miRNAs selectively sorted into HD EVs, miRNAs in blue represent non-selectively sorted miRNAs present in both LD and HD EVs. Data plotted are from 3 independent experiments (biological replicates) for miR-122, miR-142 and miR-320a and from 2 independent experiments (biological replicates) for the rest; each independent experiment with triplicate qPCR reactions; error bars represent standard deviation from independent samples. (B) TGIRT-seq results for EV miRNAs secreted by WT and La depleted cells.

miRNAs were ranked by abundance in WT-derived EVs. Their abundance was compared to those derived from La-depleted cells. MiRNAs also detected by RT-qPCR as in A are highlighted with arrows. (C) TGIRT-Seq for EVs derived from WT and La depleted backgrounds. Scatterplots showing the relative abundance (as reads per million (RPM)) for transcripts detected in EVs derived from WT and La depleted cells. Normalized per total number of reads. (D) TGIRT-Seq for WT and La depleted cells. Scatterplots showing the relative abundance (as reads per million (RPM)) for transcripts detected in cellular lysates from WT and La depleted cells. Normalized per total number of reads. The La protein gene (SSB) is also highlighted in pink to confirm La depletion by CRISPRi. Statistical significance was performed using Student's t-test (*$p<0.05$, **$p<0.01$, ns = not significant).

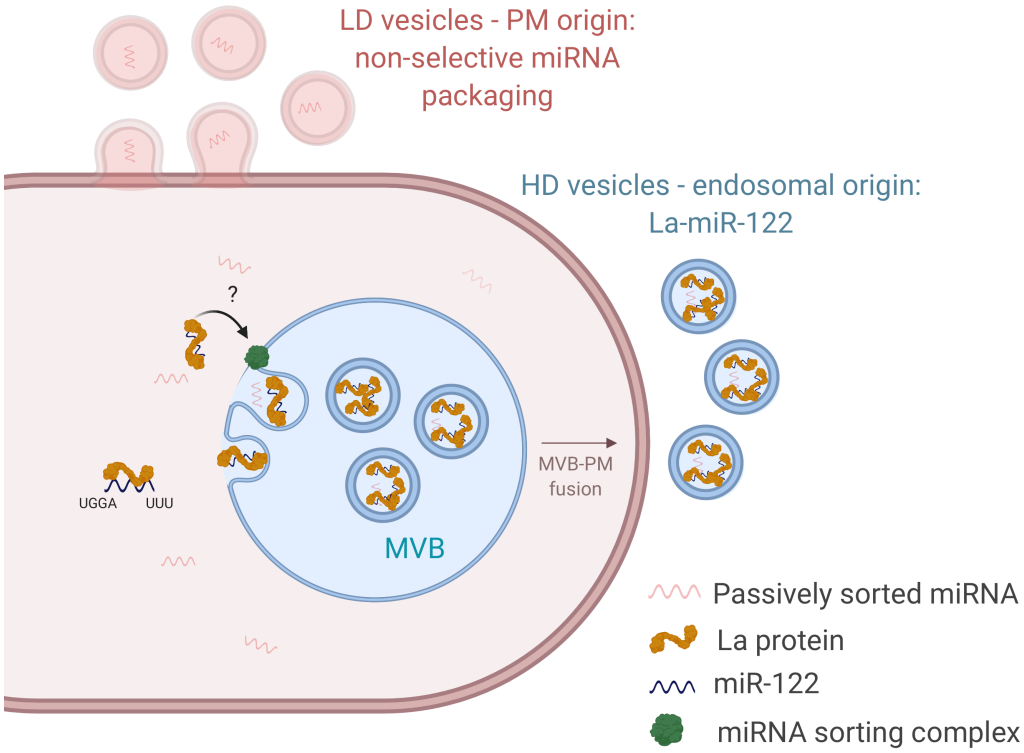


Figure 3-8: Diagram representing the current model for miR-122 secretion through exosomes derived from MDA-MB-231 cells.

Selective miRNA sorting occurs in the site of HD biogenesis (endosomal membrane). The La protein binds to selectively sorted miRNA miR-122. The La-miR-122 high affinity interaction is mediated by at least two motifs found in the 5' and 3' of miR-122. The La-miR-122 complex is recognized and sorted into intraluminal vesicles in the interior of endosomes possibly through a miRNA sorting complex in the endosomal membrane. The nature of the players in this miRNA sorting complex is investigated in *Chapter 4*. The La-miR-122 complex is then secreted as exosomal cargo. MVB: multivesicular body, PM: plasma membrane.

Chapter 4: A possible route for the EV-mediated secretion of La

Introduction

Exosomes, extracellular vesicles of endosomal origin, contain RNA and protein as soluble cargo. How the soluble cargo is sorted into exosomes remains poorly understood. Two main routes of exosome biogenesis/cargo selection have been shown. The first involves the endosomal sorting complex required for transport (ESCRT) pathway. This mechanism is initiated upon recognition and sequestration of ubiquitinated cargo. The recognition and sequestration of the cargo then leads to downstream processes that promote vesicle budding and ultimately vesicle abscission (*13*). The second route of exosome biogenesis is less well understood and does not require the ESCRT machinery. This process requires lipids, specifically ceramide, for vesicle budding and abscission (*14*). How cargo is recognized for sorting in the ceramide-dependent pathway however is still unknown.

In *Chapter 3* I found that the La protein binds specifically to miR-122 and the La-miR-122 complex gets sorted and secreted into HD vesicles. HD vesicles represent bona fide exosomes, likely derived from the endosome pathway, as shown in *Chapter 2*. In this chapter I study the possible route of La-miR-122 secretion into exosomes.

By using standard and super-resolution fluorescence microscopy I confirmed the interaction of the La protein with endosomal compartments intracellularly. The endosomal proteins potentially recognizing La for sorting into ILVs were detected by proximity labeling. I show that most of the La protein is recognized for sorting by the sequestosome-1 protein. La-containing exosomes originate through the ceramide-dependent biogenesis pathway.

Materials and Methods

Cell lines, media and general chemicals

MDA-MB-231 and HEK293T cells were cultured in DMEM with 10% FBS (Thermo Fisher Scientific, Waltham, MA). MDA-MB-231 cells were confirmed by short tandem repeat profiling (STR) and tested negative for mycoplasma contamination. For EV production, MDA-MB-231 cells were seeded at ~10% confluency in 10 % exosome-depleted FBS (System Biosciences, Palo Alto, CA) in 150mm CellBIND tissue culture dishes (Corning, Corning, NY) containing 30ml of growth medium. When HEK293T cells were used for EV collection they were seeded at ~10% confluency in FBS-free medium (similarly as the methodology used by (*15*)). EVs were collected when cells reached ~80% confluency (~72h). Unless otherwise noted, all chemicals were purchased from Sigma Aldrich (St. Louis, MO).

Extracellular vesicle purification

Conditioned medium (210ml) was harvested from 80% confluent MDA-MB-231 and HEK293T cultured cells. All subsequent manipulations were performed at 4°C. Cells

and large debris were removed by centrifugation in a Sorvall R6+ centrifuge (Thermo Fisher Scientific) at 1,000xg for 15 min followed by 10,000xg for 15 min in 500 ml vessels using a fixed angle FIBERlite F14-6X500y rotor (Thermo Fisher Scientific). The supernatant fraction was then centrifuged at ~100,000xg (28,000 RPM) using SW-28 rotors. Pellet fractions resuspended in buffer C were pooled and centrifuged at ~160,000xg (36,000 RPM) in a SW55 rotor for 1 h. The pellet was resuspended in buffer C for proteinase K assays or resuspended in 1X Laemmi buffer.

For proteinase K protection assays, the pellet resuspended in buffer C was split into 3 equal aliquots. One sample was left untreated, another sample treated with 5ug/ml proteinase K on ice for 20 min, and the last was mixed with TX-100 to a final concentration of 1% prior to proteinase K treatment. Proteinase K was inactivated with 5mM phenylmethane sulfonyl fluoride (PMSF) (5 min on ice) and samples were then mixed with Laemmli buffer for immunoblot analysis.

Construction of stable APEX2 fusion proteins

MDA-MB-231 cells expressing La-APEX2 (N), La-APEX2(C) and APEX2 alone were generated using lentiviral transduction. The La open reading frame was amplified by PCR from the plasmid pLJM60-Ssb (Addgene catalog number 38241) including or removing the stop codon depending on whether the fusion to APEX2 was placed at the coding sequence for the N- or C- terminus, respectively. For the construct La-APEX2-N, NotI and EcoRI restriction sites were added in the primers required to amplify La. For the construct La-APEX2-C, SalI and NotI restriction sites were added in the primers required to amplify La. The APEX2 open reading frame was amplified by PCR from the plasmid pcDNA3 APEX2-NES (Addgene catalog number 49386). For the construct La-APEX2-N, AgeI and NotI restriction sites were included in the primers required to amplify APEX2. For the construct La-APEX2-C, NotI and XbaI restriction sites were included in the primers required to amplify APEX2. For the construct APEX2-only, AgeI and EcoRI restriction sites were included in the primers required to amplify APEX2. The lentiviral backbone was derived from the plasmid pLJM1-EGFP (Addgene catalog number 19319). The plasmid pLJM1-EGFP was modified by removing the EGFP gene and adding the following restriction enzymes in italics as its replacement: AgeI - *SalI, NdeI, XbaI, PstI* – EcoRI. Three-way or two-way ligation was performed and the ligated product was transformed in Stbl3 competent cells.

MDA-MB-231 cells were transduced with lentivirus expressing the constructs La-APEX2-N, La-APEX2-C or APEX2 only. Cells successfully expressing the constructs were selected by adding 2ug/ml puromycin for three consecutive days. Expression of the constructs was tested by immunoblot.

***In vivo* protein biotinylation and identification of biotinylated proteins**

In vivo biotinylation was performed by following the protocol from (*151*). Briefly, MDA-MB-231 cells expressing APEX2-N, La-APEX2-C or APEX2 only were grown to 80% confluency in 4x15cm dishes. 30 ml of 500 uM biotin phenol was added to all plates other than the negative control. Cells were incubated for 30 min and immediately after H_20_2 was added to 1mM final concentration. The plates were incubated for 30 s after

which the reaction was immediately quenched with quencher solution (10 mM sodium ascorbate, 5 mM Trolox and 10 mM sodium azide solution in DPBS). Cells were washed with quencher solution 3 times and were lysed with RIPA buffer. Streptavidin beads were added to each condition and samples were incubated overnight with constant rotation at 4°C. Subsequently, the streptavidin beads were washed in 2xRIPA buffer, 1x1M KCl, 1x0.1M Na2CO3, 1x2M urea in 20 mM HEPES pH of 8, and 2x again with RIPA buffer. Proteins were eluted using elution buffer (2X Laemmli buffer supplemented with 2 mM biotin and 10 mM DTT) coupled to heating at 95°C for 10 min.

Eluted proteins were electrophoresed in a 4–20% acrylamide Tris-Glycine gradient gel (Life Technologies) for ~3 min. The bulk of proteins were stained with Coomassie and the stained band was excised from the gel using a fresh razor blade. Samples were submitted to the Taplin Mass Spectrometry Facility at Harvard University for in-gel tryptic digestion of proteins followed by liquid chromatography and mass spectrometry analysis according to their standards. The list of detected proteins was then analyzed for sub-cellular localization using the GoTermFinder developed at the Lewis-Sigler Institute at Princeton (*92*) followed by REVIGO analysis (*93*). Detected endosomal proteins were then analyzed manually in the search for proteins involved in cargo recognition.

Cellular fractionation

Fractionation of cells and membranes was done as in (*40*) with some modifications as indicated. MDA-MB-231 cells were harvested at ~80% confluency by adding cold PBS and physically removing the cells by the use of a cell scrapper. Cells were then centrifuged at 1,000xg for 10min at 4°C and cell pellets were frozen at -80°C until use. Cells were thawed and resuspended in 2 volumes of HB buffer (20mM HEPES pH 7.4, 250mM sorbitol) containing protease inhibitor cocktail 1 (1mM 4 aminobenzamidine dihydrochloride, 1 mg/ml antipain dihydrochloride, 1 mg/ml aprotinin, 1 mg/ml leupeptin, 1 mg/ml chymostatin, 1 mM phenymethylsulfonyl fluoride, 50 mM N-tosyl-L-phenylalanine chloromethyl ketone and 1 mg/ml pepstatin). Cells were passed 21-25 times through a 22 gauge needle until >80% of cells were lysed as assessed by microscopy and trypan blue staining. All steps from hereon were done on ice at 4°C, unless otherwise specified. The homogenized cells were centrifuged at 1,500xg for 30 min. The sedimented material was stored for use as "1,500g pellet" and the supernatant fraction was subsequently centrifuged at 25,000g in a TLA-55 rotor and Optima Max XP ultracentrifuge (Beckman Coulter). The sedimented material was stored to use as "25,000g pellet" and the supernatant fraction was subsequently centrifuged at 50,000g in a TLA-55 rotor. The sedimented material was stored to use as "50,000g pellet" and the supernatant fraction was subsequently centrifuged at 100,000g in a TLA-55 rotor. The sedimented material was stored to use as "100,000g pellet" and the supernatant fraction was stored to use as "100,000g sup". The pellet and supernatant fractions were mixed with Laemmli buffer and analyzed by immunoblots.

CRISPR interference

MDA-MB-231 cells expressing dCas9-KRAB, as in (*121*), were generated using lentivirus. A modified version of the transfer plasmid, as in (*121*), UCOE- EF1α-dCas9-BFP-KRAB, was kindly provided by Jonathan Weissman (UCSF). Cells were bulk

sorted for blue fluorescence protein (BFP) signal 3 d post transduction and selected cells were expanded by growth for a few generations, and then frozen and stored as parental cells (these cells are referred to as WT throughout the manuscript). Sequences for gRNAs targeting the promoter of the genes of interest were extracted from (*122*). gRNAs were cloned in plasmid pu6-sgRNA EF1Alpha-puro-T2A-BFP (*123*), plasmid #60955 obtained from Addgene, following the cloning protocol as in (*123*). The 3 top gRNAs from the V.2 library (*122*) were chosen per gene of interest. Lentiviruses with the gRNAs targeting the genes of interest were use to transduce the parental cells. Three days post transduction, cells were selected with 2ug/ml puromycin for 3 d. Post puromycin selection, cells were collected for up to 3 generations (~ 72 h) for best levels of protein depletion. More doubling times showed reduced protein depletion.

Immunoblots

Exosomes were prepared by mixing sedimented vesicles with 1X Laemmli buffer. Cell lysates were prepared by adding lysis buffer (10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 0.1% sodium dodecyl sulfate, 0.5% sodium deoxycholate, 1% Triton X-100, 10% glycerol) to cell pellets. Protein was quantified using a BCA Protein Assay Kit (Thermo Fischer Scientific), and the selected amount was mixed with Laemmli buffer. Samples were heated at 95°C for 5 min and separated on 4-20% acrylamide Tris-Glycine gradient gels (Life Technologies). Proteins were transferred to polyvinylidene difluoride membranes (EMD Millipore, Darmstadt, Germany), blocked with 5% bovine serum albumin in TBST and incubated overnight with primary antibodies. Blots were then washed with TBST, incubated with anti-rabbit or anti-mouse secondary antibodies (GE Healthcare Life Sciences, Pittsbugh, PA) and detected with ECL-2 reagent (Thermo Fisher Scientific). Primary antibodies used in this study were anti-CD9 #13174S (Cell Signaling Technology, Danvers, MA), anti-flotillin-2 #610383 (BD Biosciences, San Jose, CA), anti-Lupus La #TA500406 (Origene Technologies, Rockville, MD), anti-vinculin #ab129002 (Abcam), anti APEX #PA5-72607 (Thermo Fischer Scientific), anti VPS35 #ab10099 (Abcam), SNX17 #ab223046 (Abcam), SQSTM1 #PM045 (MBL life science), HNRNPK #ab39975 (Abcam), ELAVL #sc-5261 (Santa Cruz), SND1 #ab65078 (Abcam), STAU #ab73478 (Abcam), Histone H3 #ab201456 (Abcam), TFR #136800 (Thermo Fischer Scientific), alix #s sc-53540 (santa cruz)

Immunofluorescence

MDA-MB-231 cells on 12mm round coverslips (Corning) were fixed by adding 4% EM-grade formaldehyde (Electron Microscopy Sciences, Hatfield, PA) for 20 min at room temperature. Subsequently, cells were washed 3 times with PBS and permeabilized/blocked by adding blocking buffer (0.1% TX-100 in 2 FBS% for La validation experiments, or 0.02% saponin in 2% FBS for super-resolution structured illumination microscopy for La and Rab7 analysis (SIM)) for 20 min. Cells treated with 0.02% saponin retain intact nuclear envelopes and endoplasmic reticulum membranes (*152, 153*), which allowed us to visualize cytoplasmic La. Cells were then incubated with 1:100 dilution of anti-La (Origene Technologies, #TA500406), 1:100 dilution of anti-Rab7 (Santa Cruz, sc-6563), in blocking buffer for 1.5 h at room temperature, extensively washed and incubated in secondary antibodies diluted 1:1,000 in blocking buffer and Alexa Fluor 488 and 647 (Thermo Fischer Scientific), for 1.5 h. Cells were extensively

washed, rinsed briefly in dH_20 and mounted on slides with ProLong Gold with DAPI (Thermo Fischer). For validation experiments, WT and La-depleted cells were imaged keeping all the settings constant in an Axio Observer Z1 (Zeiss, Oberkochen, Germany). For super-resolution microscopy, images were taken with an Elyra P.1 (Zeiss)

Results

Detection of cytoplasmic La protein

In *Chapter 3,* I demonstrated that the La protein is required for the secretion of HD resident miR-122. Moreover the La protein itself co-fractionates in the HD sub-population. In *Chapter 2,* I showed evidence supporting the exosomal nature of HD vesicles. In order to confirm that the La protein is recognized and sorted into an endosomal compartment, I first examined the intracellular distribution of La. At steady-state, La was reported to be largely confined to the nucleus (*134*). However, other evidence supports a role for La function in the cytoplasm and its shuttling between the nucleus and cytoplasm (*135, 136, 138, 139*). My observation of La as an EV resident protein (Figure 3-4F) is consistent with at least some residence time in the cytoplasm.

Endogenous La was visualized by standard and structured illumination fluorescence microscopy (SIM) in fixed MDA-MB-231 cells. Immunofluorescence confirmed the presence of nuclear and cytoplasmic La (Figure 4-1A). Cells permeabilized with 0.1% Triton showed a strong nuclear signal, however the presence of cytoplasmic signal was also noted (Figure 4-1A). Triton at 0.1% concentration permeabilizes all intracellular membranes without discrimination. In order to improve the detection of cytoplasmic La I permeabilized the membranes with 0.02% Saponin. This condition allows the permeabilization of plasma membrane but fails to permeabilize the reticulum endoplasmic and the contiguous nuclear envelope (*152, 153*). By using 0.02% Saponin I successfully eliminated the La nuclear signal which permitted a better detection of cytoplasmic La (Figure 4-1B). In order to observe cytoplasmic La at a higher resolution without the nuclear signal, I used SIM with cells permeabilized with 0.02% Saponin. I observed that most of cytoplasmic La formed puncta structures smaller than 100nm of diameter (Figure 4-1C). The specificity of the antibody was affirmed by a significant depletion of the nuclear and cytoplasmic immunofluorescence signal in cells depleted of La by CRISPRi and permeabilized with 0.1% Triton (Figure 4-1D, E). Thus La is present in both, the nucleus and cytoplasm of MDA-MB-231 cells.

The cytoplasmic La protein associates with late endosomal marker Rab7

I then analyzed La co-localization with the late endosomal marker Rab7. I observed that most of the La puncta were smaller than ~100nm, however when bigger La structures were present (>500nm), they could often be observed co-localizing with Rab7 (Figure 4-2,A). This co-localization was further explored at a higher level of resolution, by using SIM (Figure 4-2,B). Thus, standard fluorescence microscopy and SIM revealed the presence of cytoplasmic puncta of La, some of which co-localized with the late endosomal marker, Rab7. Moreover these findings were confirmed by cellular fractionation. We found that a significant portion of intracellular La protein fractionates with membranes sedimented at 25,000g. This membrane fraction contains EV markers such as flotillin-2 and transferin receptor (Figure 4-2C). The La protein co-fractionating

at 25,000g might represent a pool of La tightly associated with endosomal membranes prior to sorting and/or La protein already engulfed in ILVs residing inside MVBs.

La-APEX fusion proteins are functional and secreted by EVs

Upon confirmation of the interaction between the La protein and endosomal compartments intracellularly, I sought to determine the protein(s) involved in recognizing and/or mediating La sorting into ILVs in late endosomes. I decided to use an unbiased approach consisting of proximity labeling with La-APEX2 fusion proteins. APEX2 is an evolved soybean ascorbate peroxidase able to biotinylate proteins in its close proximity upon the addition of suitable substrates. APEX2 uses H_20_2 to oxidize biotin-phenol; the biotin-phenol radical then covalently tags proteins in close proximity (*151*). By creating APEX2 fusion proteins many investigators have been able to elucidate the proteomes (*154-157*) and interactomes (*158-161*) of multiple compartments and proteins respectively. Thus we introduced constructs expressing La-APEX2 (N- and C-terminally tagged: La-APEX2-N and La-APEX2-C respectively) and cytosolic APEX2 only in WT MDA-MB-231 cells (Figure 4-3A). The three constructs showed APEX2 active enzymatic activity. Biotinylation of proteins was confirmed by streptavidin blotting (Figure 4-3B). Using the C-terminally tagged La-APEX2 fusion protein (La-APEX2-C), we found that the peroxidase activity was dependent on biotin-phenol and H_20_2 as shown by (*151*) (Figure 4-3C).

We then tested whether the La-APEX2 fusion proteins were secreted in EVs. We found that both La fusion proteins, N- and C- terminally tagged, were secreted as EV residents. La-APEX2-N and La-APEX2-C were protected from degradation by Proteinase K in the absence of non-ionic detergent, but rendered susceptible upon its addition (Figure 4-3D). This finding mirrors what was observed for endogenous secreted La – lower molecular bands in immunoblots (Figure 4-3D). Interestingly, I noticed that the C-terminally tagged protein is secreted at higher efficiency than the N-terminally tagged protein (Figure 4-3D), despite showing similar levels of intracellular expression (Figure 4-3E). I analyzed the intracellular localization of La-APEX2-N and La-APEX2-C by immunofluorescence microscopy (data not shown) and did not see any significant change in their intracellular localization. The N-terminal domains of the La protein might play a role in their recognition and sorting. I failed to detect secretion of the APEX2 enzyme alone in the high-speed crude pellet (data not shown). Thus, I concluded that the La-APEX2 fusion proteins are enzymatically active (peroxidase activity) and are secreted into EVs. Thus this system is suitable for studying possible La interactors mediating La sorting into EVs.

The La-APEX2-C fusion protein helps to define possible routes for La secretion

Since La-APEX2-C was secreted at higher efficiency than the N-terminally tagged counterpart (Figure 4-3D) we decided to use this fusion protein for further investigation. MDA-MB-231 cells expressing either La-APEX2-C or cytosolic APEX2 alone were exposed to biotin-phenol for 30 min. Upon incubation with biotin-phenol, I exposed the cells to H_20_2 for 30 sec. The peroxidase activity was quenched immediately after. Cells were lysed with RIPA buffer and biotinylated proteins were captured with streptavidin beads. Upon binding, stringent washes were performed and bound proteins were eluted

with Laemmli buffer supplemented with biotin. Samples were then heated to 95°C for 15 min. Bound proteins were analyzed by mass spectrometry. Only proteins with at least 2 unique peptides were considered for further analysis. Proteins enriched at least by 2-fold as the ratio La-APEX-2-C/APEX2 were considered for further analysis. Gene ontology analysis of these proteins showed that terms such as "exosome and endosome" were enriched (Figure 4-4A). Thus confirming the association of the La protein with endosomes and its exosomal secretion. The La-APEX2-C fusion protein had mostly a cytoplasmic localization (a nuclear exit signal is included in the APEX2 ORF), thus a nuclear localization is absent in the GO analysis of biotinylated proteins. I then curated the data for only endosome-associated proteins. From this list, I focused on those that have been previously reported to recognize cargo for sorting for different biological processes. By doing so, I shortened the list of potential candidates mediating La recognition and sorting to three proteins: vacuolar protein sorting-associated protein 35 (VPS35), sorting nexin-17 (SNX17) and sequestosome-1 (SQSTM1) (Figure 4-4B). VPS35 and SNX17 have been previously shown to mediate cargo sorting in the retromer complex (*162, 163*) and SQSTM1 to recognize cargo in the autophagy pathway (*164*).

I used CRISPRi to knock down the three La sorting candidates (Figure 4-4C). Extracellular vesicles derived from WT cells or cells depleted for VPS35, SNX17 or SQSTM1 were isolated. La secretion was normalized to number of cells and evaluated by La immunoblots. Similar number of cells was seeded at the beginning of the experiment; their number remained comparable throughout the experiment. We observed a dramatic effect on La secretion upon SQSTM1 depletion (Figure 4-4D). These experiments suggest a role of SQSTM1 in La sorting into exosomes.

The La protein is secreted by an LC3-dependent extracellular vesicle loading and secretion

A recent study has shown that a subset of lipidated LC3 (LC3-II) localizes to late endosomes. The endosomal LC3 is required for the recruitment of the neutral sphingomyelinase SMPD3. SMPD3 promotes EV biogenesis via the ceramide-dependent biogenesis pathway (*15*). Our preliminary results showed the requirement of SQSTM1 for La secretion. SQSTM1 bears domains that bind to LC3 and ubiquitinated proteins simultaneously (*164*). Thus this protein is known to facilitate the capture of ubiquitinated cargoes to LC3 containing autophagosomes, therefore recognizing and sorting cargoes to be degraded via the autophagy pathway. The requirement of SQSTM1 for EV-mediated La secretion (Figure 4-4D) and the recent findings on LC3 requirement for EV biogenesis led us to hypothesize that La sorting is dependent on the newly discovered LC3-dependent pathway.

Leidal et al. (2020) showed that by knocking out ATG7, a protein required for LC3 lipidation, the secretion of RBPs hnRNPK and SAFB was impaired (*15*). WT and ATG7 -/- HEK293T cells were obtained from the aforementioned study and La secretion was tested by immunoblots. I reproduced Leidal et al. (2020) findings in respect to the requirement of ATG7 for hnRNPK secretion (Figure 4-4A, B). Importantly, I found that La secretion was dependent on ATG7 (Figure 4-4A, B). ELAVL1 is a RBP reported to be required for miR-122 secretion in hepatocytes (*128*). I found that ELAVL1 is a strong La

interactor as shown by La co-IP (Figure 4-4C). Notably ELAVL secretion was also dependent on ATG7 (Figure 4-4A, B). Other EV-present RBPs such as SND1 and STAU do not require ATG7 (Figure 4-4A, B). Thus the La protein and its strong interactor ELAVL1 require ATG7, and LC3 lipidation, for their secretion.

My results are suggestive of an LC3-dependent mechanism of La sorting. The La protein may be recognized by SQSTM1 and recruited to LC3-containing endosomes. LC3 then or concomitantly, recruits SMPD3 and promotes ceramide-dependent EV biogenesis. This model presumes that the La protein targeted for recognition and secretion must be ubiquitinated. Whether the La protein is post-translationally modified with ubiquitin or a ubiquitin-like molecule prior to sorting into ILVs requires further work.

Discussion

Here, I document a possible route for La recognition and secretion into EVs. By using proximity labeling and genetic tools, I found that the La protein requires SQSTM1 and lipidated LC3 (LC3-II) for its secretion (Figure 4-6).

The La protein has been reported to largely confine to the nucleus at steady state (*134*). However, other evidence supports a role for La function in the cytoplasm and its shuttling between the nucleus and cytoplasm (*135, 136, 138, 139*). I probed the cytoplasmic La pool by immunofluorescence microscopy. I found that the cytoplasmic La is largely present as punctate structures (Figure 4-1C), some of which closely associate with the late endosomal marker Rab7 (Figure 4-2B). Interestingly, I observed that the La puncta associated with Rab7 structures showed a peripheral association to the endosomes (Figure 4-2B). This peripheral association suggests that I captured the moment prior to La sorting into ILVs. La-containing aggregates or RNA granules might be recruited to Rab7 positive endosomes through SQSTM1. Alternatively these granules might be tethered to endosomes through annexins, as shown for other RNA granules and lysosomes (*58*). Their recognition and sorting into ILVs might occur at a later step post annexin-mediated tethering. Studying the involvement of annexins in these steps will be informative.

I used proximity labeling as an unbiased method to determine what endosomal proteins might mediate La recognition and sorting into ILVs. By using this technique and CRISPRi to deplete the candidate proteins, I determined that SQSTM1 is required for La secretion. Importantly SQSTM1 has been previously detected in ILVs residing in late endosomes (*165*) and its presence in the high-speed pellet fraction from conditioned medium from multiple cell lines has been reported (*62, 130, 166, 167*). SQSTM1 recognizes its cargo by binding to ubiquitinated proteins (*164*). We have not consistently detected ubiquitinated forms of the La protein in EVs. One possibility is that deubiquitinating enzymes (DUB) act on La as it enters a nascent ILV on the surface of an endosome. I have detected the presence of DUBs in HD and LD EVs derived from MDA-MB-231 by mass spectrometry (data not shown). Thus it is possible that ubiquitinated La is recognized by SQSTM1 for its secretion, but the detection of ubiquitinated La is hampered by the action of DUBs present in EVs. Mutating potential sites for ubiquitinylation in the La protein would help answer this question.

A recent study has discovered a new pathway for exosomal biogenesis and loading dependent on LC3 (*15*). This new pathway is termed LC3-dependent extracellular vesicle loading and secretion (LDELS). I found that La secretion is dependent on this pathway (Figure 4-5A, B). This finding coupled to our previous results on the requirement of SQSTM1for La secretion (Figure 4-4D) fit well and lead us to propose a model for La recognition and secretion. We propose that the La protein destined for secretion forms moderate size aggregates/granules and that some or all La species in these aggregates might be ubiquitinated. Ubiquitinylated proteins are recognized by SQSTM1. SQSTM1, through its LC-3 binding domain, then binds to lipidated LC3 in the endosomal membrane and recruits the ubiquitinated La aggregates/granules to the site of exosome biogenesis thus promoting its sorting. What other RNA binding proteins co-localize to the La aggregates/granules and what their functions are will be informative. It remains to be seen if ubiquitin or other post-translational modifications of La promote its sorting into nascent exosomes.

Figures

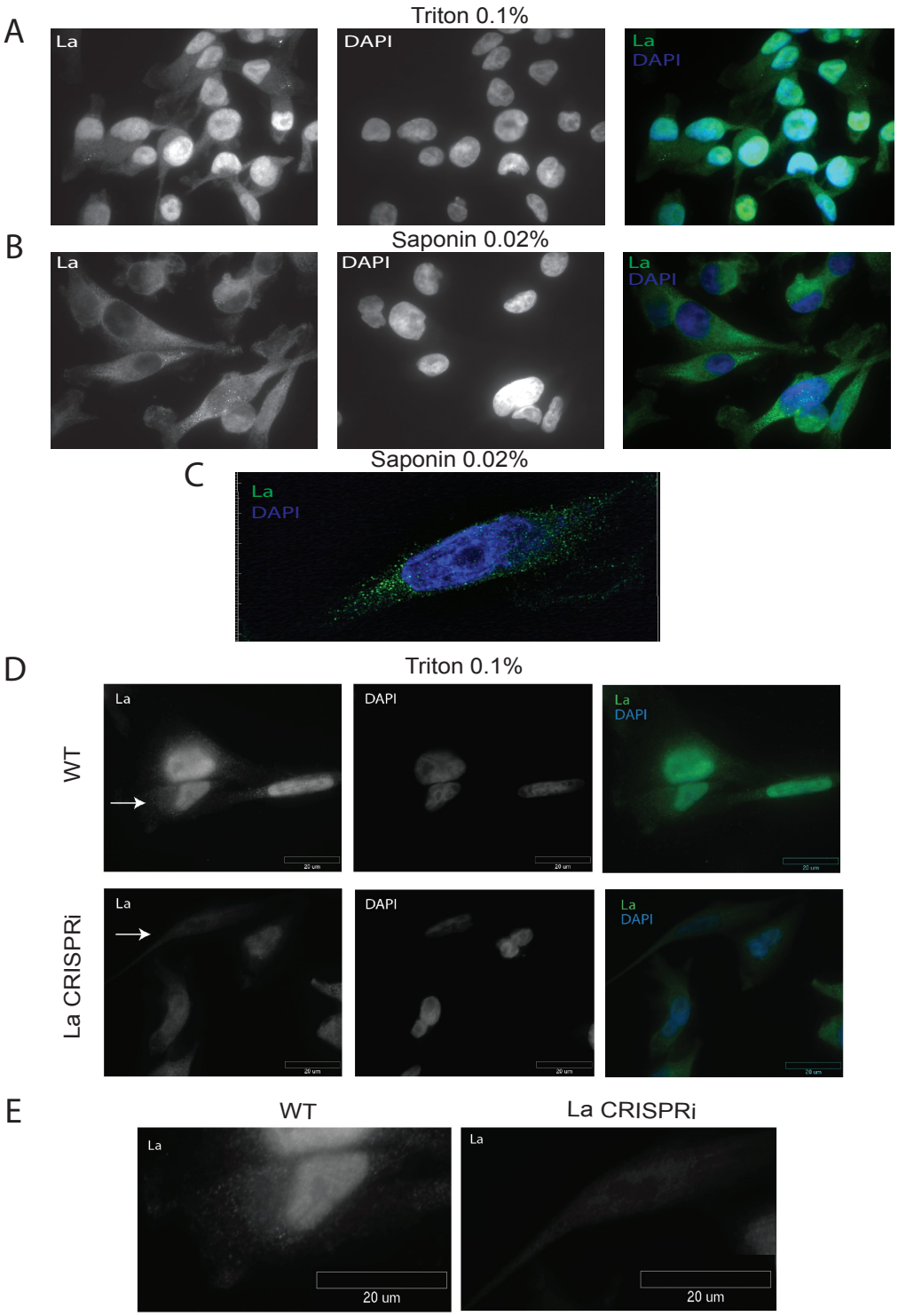


Figure 4-1: Detection of nuclear and cytoplasmic La by immunofluorescence.

(A) Immunofluorescence for endogenous La was performed. WT cells were permeabilized with 0.1% Triton X-100. La and DAPI staining are shown. (B) Immunofluorescence for endogenous La was performed. WT cells were permeabilized with 0.02% Saponin. La and DAPI staining are shown. (C) WT cells were treated as in B.

Structured illumination microscopy (SIM) was used to achieve higher resolution. (D) Antibody specificity was confirmed by standard fluorescence microscopy in WT and La-depleted cells. WT and La-depleted cells were fixed and permeabilized with 0.1% Triton X-100 as in A. La and DAPI staining are shown. (E) Cells with arrows in D are shown side by side to confirm antibody specificity.

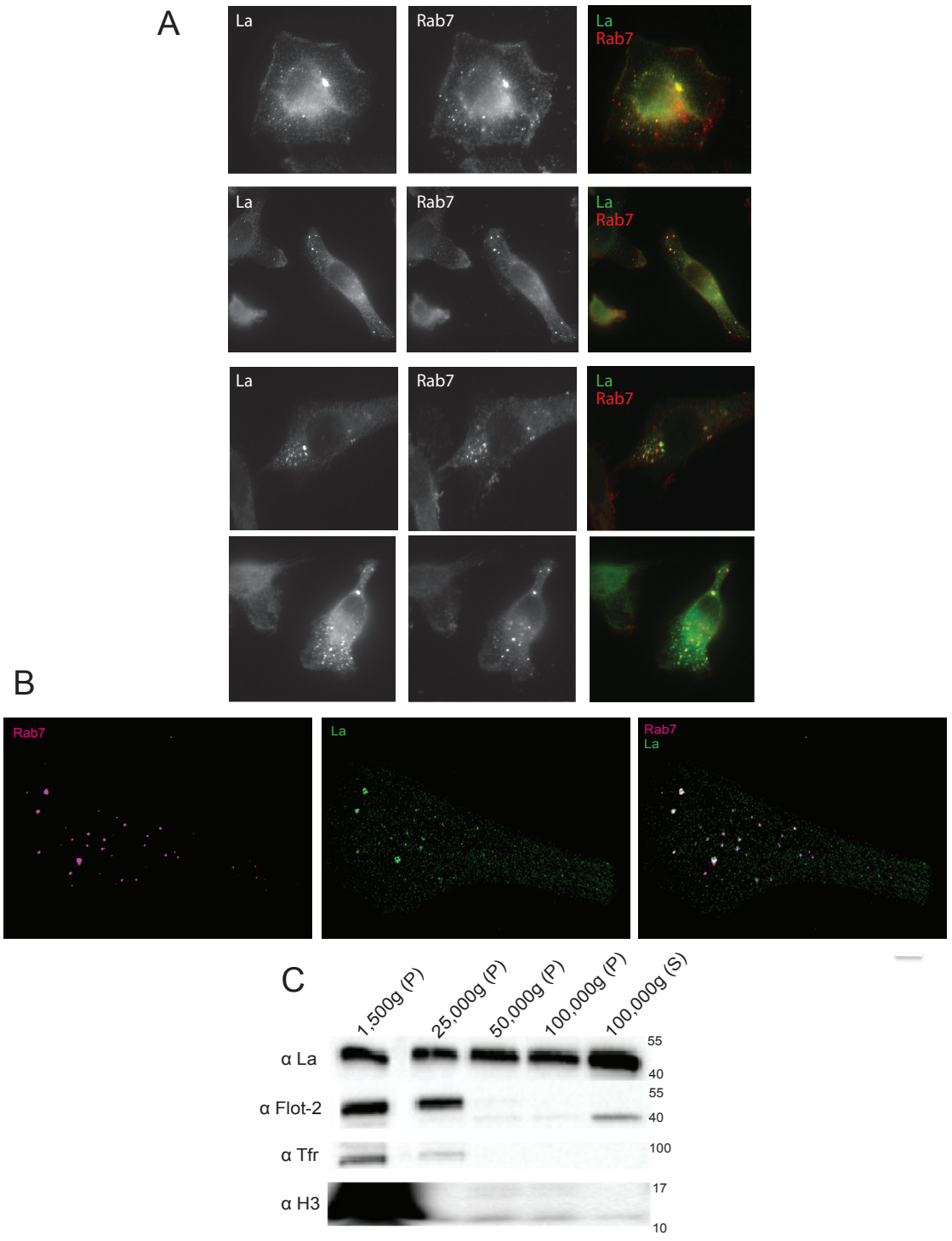


Figure 4-2: La associates with Rab7-positive vesicles.

(A) Immunofluorescence for endogenous La and Rab7 was performed. WT cells were fixed and permeabilized with 0.02% saponin in order to preserve the nuclear envelope intact and avoid detection of nuclear La. (B) Structured illumination microscopy (SIM) was used to achieve higher resolution. (C) Cellular fractionation. The

La protein fractionates with various particulate fractions. Immunoblots for La, flotillin-2, Tfr and histone 3 are shown. P: pellet, S: supernatant. Tfr: transferrin receptor.

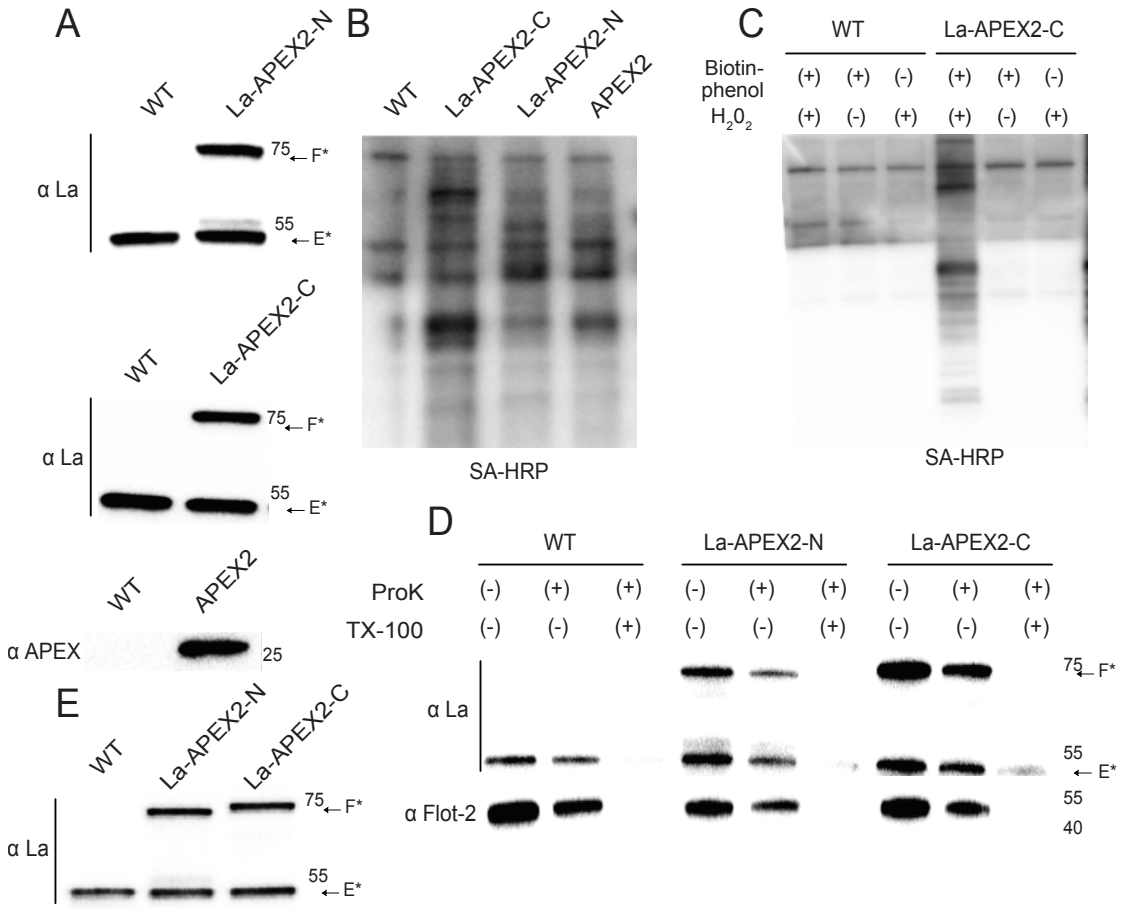


Figure 4-3: Functionality and secretion of La-APEX2 fusion proteins.

(A) Immunoblots to confirm the expression of the APEX2 constructs: La-APEX2-N, La-APEX2-C and APEX2. Immunoblots for La and APEX2 are shown. (B) Confirmation of APEX2 peroxidase activity in the three different constructs. Biotinylated proteins were detected by streptavidin blotting. (C) Characterization of APEX2-mediated biotinylation requirements. Cells were incubated in the presence or absence of biotin-phenol or H_20_2. Biotinylated proteins were detected by streptavidin blotting. (D) Proteinase K protection assays in high-speed pellet fractions derived from WT, La-APEX2-N and La-Apex2-C cells. Samples were treated with or without proteinase K without or with Triton X-100. Immunoblots for La and flotillin-2 are shown. The La-APEX2 fusion proteins behave similarly to endogenous La (lower band in the gel). (E) Immunoblots comparing the level of expression of La-APEX2-N and La-APEX2-C. Immunoblots for La are shown. F*: fusion protein, E*: endogenous protein.

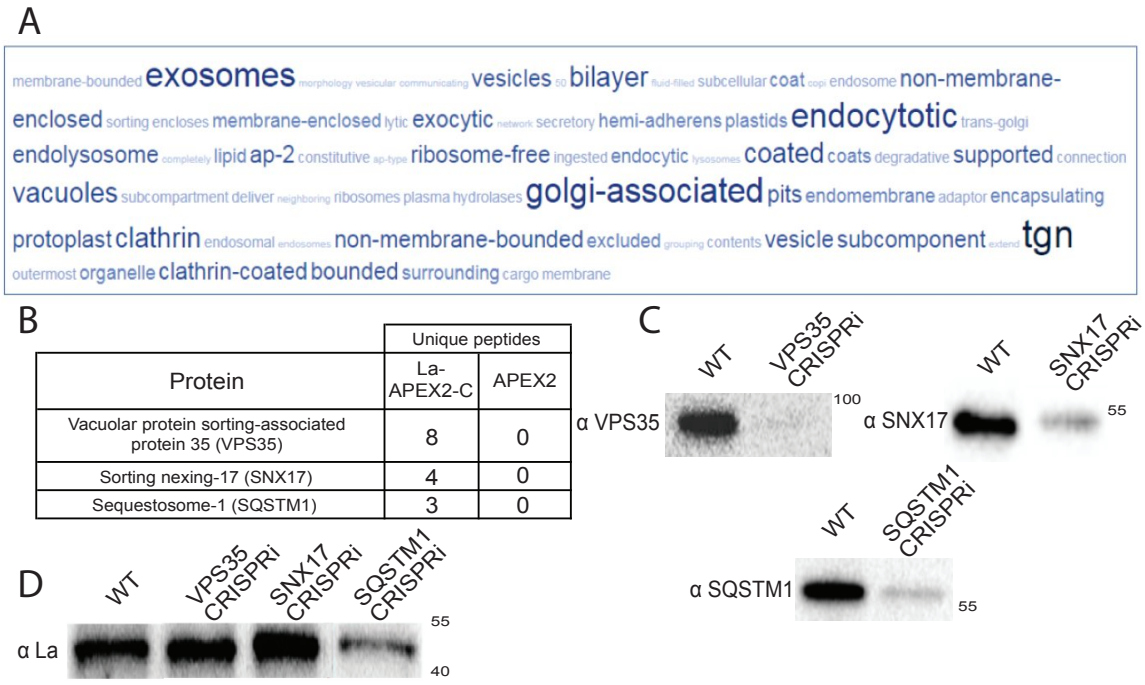


Figure 4-4: Proximity biotinylation suggests the requirement of SQSTM-1 for La secretion.

(A) Biotinylated proteins by La-APEX2-C show enrichment for endosomal and exosomal localization. Peptides detected in La-APEX2-C and APEX2 were compared to each other and only peptides enriched by 2-fold or more in La-APEX2-C were considered for Gene Ontology analysis. The gene ontology analysis compares the list of proteins given per group to the human genome frequency. The detection of enrichment for certain organelle/ localization is represented by words. The word size is correlated to its enrichment level. (B) List of proteins from A with endosomal localization and with previous record of recognizing cargo in distinct biological processes. (C) Analysis of the proteins of interest by immunoblots upon CRISPRi depletion. Immunoblots for VPS35, SNX17 and SQSTM1 are shown. Immunoblots were normalized to protein concentration; equal amounts of protein were loaded per lane. (D) Immunoblots for La secreted in EVs derived from WT, VPS35, SNX17 and SQSTM1 depleted cells are shown. La secretion was normalized to cell number. Similar number of cells was seeded at the beginning of the experiment. Their relative numbers remain similar throughout the experiment.

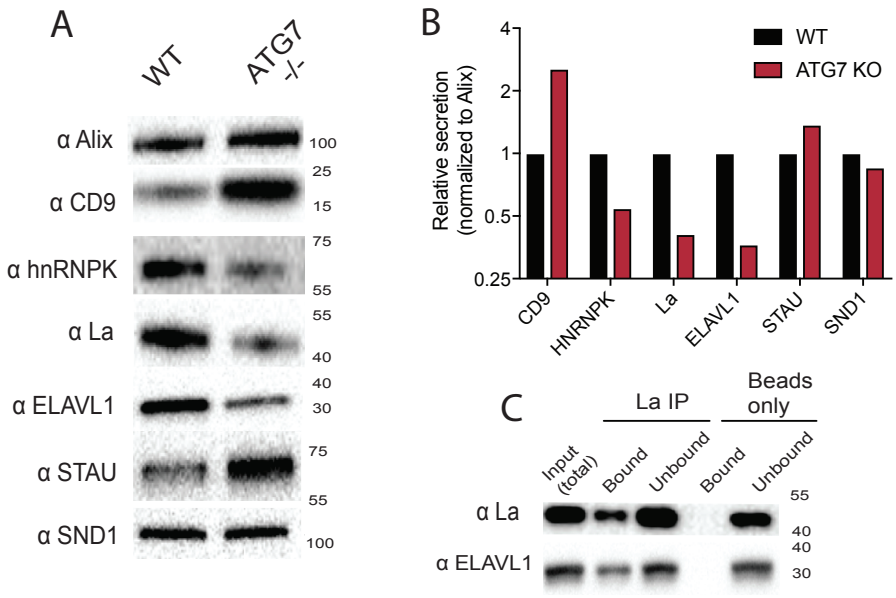


Figure 4-5: ATG7 is required for La secretion.

(A) Secretion of distinct RBPs by WT and ATG7-/- cells. Immunoblots for Alix, CD9, hnRNPK, La, ELAVL1, STAU and SND1 are shown. (B) Quantification of A. Leidal et al. (2020) showed no difference on Alix secretion upon ATG7 deletion, thus data was normalized to Alix secretion. (C) Immunoprecipitation of endogenous La. Immunoblots for La and ELAVL post-La IP are shown.

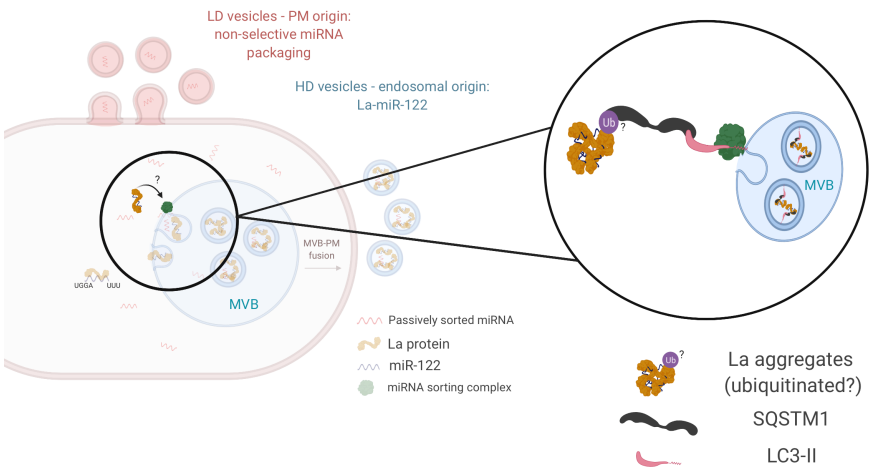


Figure 4-6: Diagram representing the current model of La recognition in the endosomal membrane for its exosomal-mediated secretion.

The La protein forms aggregates/granules prior to sorting. The La protein aggregates/granules are recognized by SQSTM1, possibly through the attachment of ubiquitin or an ubiquitin-like molecule to La (this is yet to be proven). Once the La aggregates/granules are recognized by SQSTM1 they are recruited to LC3-II containing endosomes to proceed with their sorting into intraluminal vesicles in the MVB.

References

1. T. Pisitkun, R. F. Shen, M. A. Knepper, Identification and proteomic profiling of exosomes in human urine. *Proc Natl Acad Sci U S A* **101**, 13368-13373 (2004).
2. M. P. Caby, D. Lankar, C. Vincendeau-Scherrer, G. Raposo, C. Bonnerot, Exosomal-like vesicles are present in human blood plasma. *Int Immunol* **17**, 879-887 (2005).
3. C. Admyre *et al.*, Exosomes with immune modulatory features are present in human breast milk. *J Immunol* **179**, 1969-1978 (2007).
4. L. J. Vella *et al.*, Packaging of prions into exosomes is associated with a novel pathway of PrP processing. *J Pathol* **211**, 582-590 (2007).
5. M. Colombo, G. Raposo, C. Thery, Biogenesis, secretion, and intercellular interactions of exosomes and other extracellular vesicles. *Annu Rev Cell Dev Biol* **30**, 255-289 (2014).
6. J. Kowal *et al.*, Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. *Proc Natl Acad Sci U S A* **113**, E968-977 (2016).
7. J. P. Tosar *et al.*, Assessment of small RNA sorting into different extracellular fractions revealed by high-throughput sequencing of breast cell lines. *Nucleic Acids Res* **43**, 5601-5616 (2015).
8. M. M. Temoche-Diaz *et al.*, Distinct mechanisms of microRNA sorting into cancer cell-derived extracellular vesicle subtypes. *Elife* **8**, (2019).
9. C. Harding, J. Heuser, P. Stahl, Receptor-mediated endocytosis of transferrin and recycling of the transferrin receptor in rat reticulocytes. *J Cell Biol* **97**, 329-339 (1983).
10. B. T. Pan, K. Teng, C. Wu, M. Adam, R. M. Johnstone, Electron microscopic evidence for externalization of the transferrin receptor in vesicular form in sheep reticulocytes. *J Cell Biol* **101**, 942-948 (1985).
11. E. Cocucci, G. Racchetti, J. Meldolesi, Shedding microvesicles: artefacts no more. *Trends Cell Biol* **19**, 43-51 (2009).
12. S. R. Elkin, A. M. Lakoduk, S. L. Schmid, Endocytic pathways and endosomal trafficking: a primer. *Wien Med Wochenschr* **166**, 196-204 (2016).
13. J. H. Hurley, The ESCRT complexes. *Crit Rev Biochem Mol Biol* **45**, 463-487 (2010).
14. K. Trajkovic *et al.*, Ceramide triggers budding of exosome vesicles into multivesicular endosomes. *Science* **319**, 1244-1247 (2008).
15. A. M. Leidal *et al.*, The LC3-conjugation machinery specifies the loading of RNA-binding proteins into extracellular vesicles. *Nat Cell Biol* **22**, 187-199 (2020).
16. M. F. Baietti *et al.*, Syndecan-syntenin-ALIX regulates the biogenesis of exosomes. *Nat Cell Biol* **14**, 677-685 (2012).
17. A. Savina, M. Vidal, M. I. Colombo, The exosome pathway in K562 cells is regulated by Rab11. *J Cell Sci* **115**, 2505-2515 (2002).
18. M. Ostrowski *et al.*, Rab27a and Rab27b control different steps of the exosome secretion pathway. *Nat Cell Biol* **12**, 19-30; sup pp 11-13 (2010).

19. C. Hsu *et al.*, Regulation of exosome secretion by Rab35 and its GTPase-activating proteins TBC1D10A-C. *J Cell Biol* **189**, 223-232 (2010).
20. V. Muralidharan-Chari *et al.*, ARF6-regulated shedding of tumor cell-derived plasma membrane microvesicles. *Curr Biol* **19**, 1875-1885 (2009).
21. J. F. Nabhan, R. Hu, R. S. Oh, S. N. Cohen, Q. Lu, Formation and release of arrestin domain-containing protein 1-mediated microvesicles (ARMMs) at plasma membrane by recruitment of TSG101 protein. *Proc Natl Acad Sci U S A* **109**, 4146-4151 (2012).
22. E. Willms *et al.*, Cells release subpopulations of exosomes with distinct molecular and biological properties. *Sci Rep* **6**, 22519 (2016).
23. D. K. Jeppesen *et al.*, Reassessment of Exosome Composition. *Cell* **177**, 428-445 e418 (2019).
24. B. K. Thakur *et al.*, Double-stranded DNA in exosomes: a novel biomarker in cancer detection. *Cell Res* **24**, 766-769 (2014).
25. A. Yokoi *et al.*, Mechanisms of nuclear content loading to exosomes. *Sci Adv* **5**, eaax8849 (2019).
26. M. J. Shurtleff *et al.*, Broad role for YBX1 in defining the small noncoding RNA composition of exosomes. *Proc Natl Acad Sci U S A* **114**, E8987-E8995 (2017).
27. H. Valadi *et al.*, Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol* **9**, 654-659 (2007).
28. D. P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-297 (2004).
29. M. Mittelbrunn *et al.*, Unidirectional transfer of microRNA-loaded exosomes from T cells to antigen-presenting cells. *Nat Commun* **2**, 282 (2011).
30. A. Montecalvo *et al.*, Mechanism of transfer of functional microRNAs between mouse dendritic cells via exosomes. *Blood* **119**, 756-766 (2012).
31. D. M. Pegtel *et al.*, Functional delivery of viral miRNAs via exosomes. *Proc Natl Acad Sci U S A* **107**, 6328-6333 (2010).
32. M. Y. Fong *et al.*, Breast-cancer-secreted miR-122 reprograms glucose metabolism in premetastatic niche to promote metastasis. *Nat Cell Biol* **17**, 183-194 (2015).
33. C. T. Dickman *et al.*, Selective extracellular vesicle exclusion of miR-142-3p by oral cancer cells promotes both internal and extracellular malignant phenotypes. *Oncotarget* **8**, 15252-15266 (2017).
34. W. Zhou *et al.*, Cancer-secreted miR-105 destroys vascular endothelial barriers to promote metastasis. *Cancer Cell* **25**, 501-515 (2014).
35. N. Tominaga *et al.*, Brain metastatic cancer cells release microRNA-181c-containing extracellular vesicles capable of destructing blood-brain barrier. *Nat Commun* **6**, 6716 (2015).
36. Y. L. Hsu *et al.*, Hypoxic lung cancer-secreted exosomal miR-23a increased angiogenesis and vascular permeability by targeting prolyl hydroxylase and tight junction protein ZO-1. *Oncogene* **36**, 4929-4942 (2017).
37. N. Kosaka *et al.*, Neutral sphingomyelinase 2 (nSMase2)-dependent exosomal transfer of angiogenic microRNAs regulate cancer cell metastasis. *J Biol Chem* **288**, 10849-10859 (2013).

38. N. Kosaka, Y. Yoshioka, K. Hagiwara, N. Tominaga, T. Ochiya, Functional analysis of exosomal microRNA in cell-cell communication research. *Methods Mol Biol* **1024**, 1-10 (2013).
39. N. Kosaka *et al.*, Secretory mechanisms and intercellular transfer of microRNAs in living cells. *J Biol Chem* **285**, 17442-17452 (2010).
40. M. J. Shurtleff, M. M. Temoche-Diaz, K. V. Karfilis, S. Ri, R. Schekman, Y-box protein 1 is required to sort microRNAs into exosomes in cells and in a cell-free reaction. *Elife* **5**, (2016).
41. J. Van Deun *et al.*, The impact of disparate isolation methods for extracellular vesicles on downstream RNA profiling. *J Extracell Vesicles* **3**, (2014).
42. C. Villarroya-Beltri *et al.*, Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs. *Nat Commun* **4**, 2980 (2013).
43. L. Santangelo *et al.*, The RNA-Binding Protein SYNCRIP Is a Component of the Hepatocyte Exosomal Machinery Controlling MicroRNA Sorting. *Cell Rep* **17**, 799-808 (2016).
44. D. Enderle *et al.*, Characterization of RNA from Exosomes and Other Extracellular Vesicles Isolated by a Novel Spin Column-Based Method. *PLoS One* **10**, e0136133 (2015).
45. S. A. Bellingham, B. M. Coleman, A. F. Hill, Small RNA deep sequencing reveals a distinct miRNA signature released in exosomes from prion-infected neuronal cells. *Nucleic Acids Res* **40**, 10937-10949 (2012).
46. E. N. Nolte-'t Hoen *et al.*, Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions. *Nucleic Acids Res* **40**, 9272-9285 (2012).
47. C. Lasser *et al.*, Two distinct extracellular RNA signatures released by a single cell type identified by microarray and next-generation sequencing. *RNA Biol* **14**, 58-72 (2017).
48. N. Guzman *et al.*, Breast Cancer-Specific miR Signature Unique to Extracellular Vesicles Includes "microRNA-like" tRNA Fragments. *Mol Cancer Res* **13**, 891-901 (2015).
49. J. P. Tosar *et al.*, Fragmentation of extracellular ribosomes and tRNAs shapes the extracellular RNAome. *bioRxiv*, (2020).
50. A. H. Buck *et al.*, Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity. *Nat Commun* **5**, 5488 (2014).
51. B. Y. Nabet *et al.*, Exosome RNA Unshielding Couples Stromal Activation to Pattern Recognition Receptor Signaling in Cancer. *Cell* **170**, 352-366 e313 (2017).
52. T. A. P. Driedonks *et al.*, Immune stimuli shape the small non-coding transcriptome of extracellular vesicles released by dendritic cells. *Cell Mol Life Sci* **75**, 3857-3875 (2018).
53. E. L. Garcia *et al.*, Packaging of host mY RNAs by murine leukemia virus may occur early in Y RNA biogenesis. *J Virol* **83**, 12526-12534 (2009).
54. A. A. Onafuwa-Nuga, S. R. King, A. Telesnitsky, Nonrandom packaging of host RNAs in moloney murine leukemia virus. *J Virol* **79**, 13528-13537 (2005).

55. J. Skog *et al.*, Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat Cell Biol* **10**, 1470-1476 (2008).
56. C. P. Lai *et al.*, Visualization and tracking of tumour extracellular vesicle delivery and RNA translation using multiplexed reporters. *Nat Commun* **6**, 7029 (2015).
57. O. Meyuhas, T. Kahan, The race to decipher the top secrets of TOP mRNAs. *Biochim Biophys Acta* **1849**, 801-811 (2015).
58. Y. C. Liao *et al.*, RNA Granules Hitchhike on Lysosomes for Long-Distance Transport, Using Annexin A11 as a Molecular Tether. *Cell* **179**, 147-164 e120 (2019).
59. K. Hagiwara, T. Katsuda, L. Gailhouste, N. Kosaka, T. Ochiya, Commitment of Annexin A2 in recruitment of microRNAs into extracellular vesicles. *FEBS Lett* **589**, 4071-4078 (2015).
60. S. Atay *et al.*, Oncogenic KIT-containing exosomes increase gastrointestinal stromal tumor cell invasion. *Proc Natl Acad Sci U S A* **111**, 711-716 (2014).
61. C. Fedele, A. Singh, B. J. Zerlanko, R. V. Iozzo, L. R. Languino, The alphavbeta6 integrin is transferred intercellularly via exosomes. *J Biol Chem* **290**, 4545-4551 (2015).
62. H. Peinado *et al.*, Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nat Med* **18**, 883-891 (2012).
63. S. Rana, K. Malinowska, M. Zoller, Exosomal tumor microRNA modulates premetastatic organ cells. *Neoplasia* **15**, 281-295 (2013).
64. H. C. Christianson, M. Belting, Heparan sulfate proteoglycan as a cell-surface endocytosis receptor. *Matrix Biol* **35**, 51-55 (2014).
65. A. Zomer *et al.*, In Vivo imaging reveals extracellular vesicle-mediated phenocopying of metastatic behavior. *Cell* **161**, 1046-1057 (2015).
66. H. C. Christianson, K. J. Svensson, T. H. van Kuppevelt, J. P. Li, M. Belting, Cancer cell exosomes depend on cell-surface heparan sulfate proteoglycans for their internalization and functional activity. *Proc Natl Acad Sci U S A* **110**, 17380-17385 (2013).
67. W. Heusermann *et al.*, Exosomes surf on filopodia to enter cells at endocytic hot spots, traffic within endosomes, and are targeted to the ER. *J Cell Biol* **213**, 173-184 (2016).
68. B. H. Sung *et al.*, A live cell reporter of exosome secretion and uptake reveals pathfinding behavior of migrating cells. *Nat Commun* **11**, 2092 (2020).
69. M. E. Hung, J. N. Leonard, A platform for actively loading cargo RNA to elucidate limiting steps in EV-mediated delivery. *J Extracell Vesicles* **5**, 31027 (2016).
70. K. J. Svensson *et al.*, Exosome uptake depends on ERK1/2-heat shock protein 27 signaling and lipid Raft-mediated endocytosis negatively regulated by caveolin-1. *J Biol Chem* **288**, 17713-17724 (2013).
71. J. Grove, M. Marsh, The cell biology of receptor-mediated virus entry. *J Cell Biol* **195**, 1071-1082 (2011).
72. M. Marsh, A. Helenius, Virus entry: open sesame. *Cell* **124**, 729-740 (2006).

73. P. P. Luyet, T. Falguieres, V. Pons, A. K. Pattnaik, J. Gruenberg, The ESCRT-I subunit TSG101 controls endosome-to-cytosol release of viral RNA. *Traffic* **9**, 2279-2290 (2008).
74. R. C. Piper, D. J. Katzmann, Biogenesis and function of multivesicular bodies. *Annu Rev Cell Dev Biol* **23**, 519-547 (2007).
75. A. Jegou *et al.*, CD9 tetraspanin generates fusion competent sites on the egg membrane for mammalian fertilization. *Proc Natl Acad Sci U S A* **108**, 10946-10951 (2011).
76. K. Miyado *et al.*, Requirement of CD9 on the egg plasma membrane for fertilization. *Science* **287**, 321-324 (2000).
77. M. C. Boelens *et al.*, Exosome transfer from stromal to breast cancer cells regulates therapy resistance pathways. *Cell* **159**, 499-513 (2014).
78. B. Costa-Silva *et al.*, Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nat Cell Biol* **17**, 816-826 (2015).
79. A. Hoshino *et al.*, Tumour exosome integrins determine organotropic metastasis. *Nature* **527**, 329-335 (2015).
80. M. L. Squadrito *et al.*, Endogenous RNAs modulate microRNA sorting to exosomes and transfer to acceptor cells. *Cell Rep* **8**, 1432-1446 (2014).
81. C. P. Lai *et al.*, Dynamic biodistribution of extracellular vesicles in vivo using a multimodal imaging reporter. *ACS Nano* **8**, 483-494 (2014).
82. K. E. van der Vos *et al.*, Directly visualized glioblastoma-derived extracellular vesicles transfer RNA to microglia/macrophages in the brain. *Neuro Oncol* **18**, 58-69 (2016).
83. K. Miyauchi, Y. Kim, O. Latinovic, V. Morozov, G. B. Melikyan, HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes. *Cell* **137**, 433-444 (2009).
84. B. T. Pan, R. M. Johnstone, Fate of the transferrin receptor during maturation of sheep reticulocytes in vitro: selective externalization of the receptor. *Cell* **33**, 967-978 (1983).
85. M. S. Ostenfeld *et al.*, Cellular disposal of miR23b by RAB27-dependent exosome release is linked to acquisition of metastatic properties. *Cancer Res* **74**, 5758-5771 (2014).
86. A. K. Chen *et al.*, MicroRNA binding to the HIV-1 Gag protein inhibits Gag assembly and virus production. *Proc Natl Acad Sci U S A* **111**, E2676-2683 (2014).
87. S. L. Wolin *et al.*, Non-coding Y RNAs as tethers and gates: Insights from bacteria. *RNA Biol* **10**, 1602-1608 (2013).
88. K. M. Reinisch, S. L. Wolin, Emerging themes in non-coding RNA quality control. *Curr Opin Struct Biol* **17**, 209-214 (2007).
89. J. R. Hogg, K. Collins, Human Y5 RNA specializes a Ro ribonucleoprotein for 5S ribosomal RNA quality control. *Genes Dev* **21**, 3067-3072 (2007).
90. S. Kruger *et al.*, Molecular characterization of exosome-like vesicles from breast cancer cells. *BMC Cancer* **14**, 44 (2014).
91. A. Bobrie, M. Colombo, S. Krumeich, G. Raposo, C. Thery, Diverse subpopulations of vesicles secreted by different intracellular mechanisms are

present in exosome preparations obtained by differential ultracentrifugation. *J Extracell Vesicles* **1**, (2012).
92. E. I. Boyle *et al.*, GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* **20**, 3710-3715 (2004).
93. F. Supek, M. Bosnjak, N. Skunca, T. Smuc, REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
94. Y. Qin *et al.*, High-throughput sequencing of human plasma RNA by using thermostable group II intron reverse transcriptases. *RNA*. **22**, 111-128. doi: 110.1261/rna.054809.054115. Epub 052015 Nov 054809. (2016).
95. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).
96. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360 (2015).
97. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
98. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
99. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
100. P. D. Hsu *et al.*, DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832 (2013).
101. H. Lodish *et al.*, *Molecular Cell Biology, 4th edition*. (W. H. Freeman and Company, 2000).
102. S. Gould, F. K. Fordjour, G. Daaboul, A shared pathway of exosome biogenesis operates at plasma and endosome membranes. *bioRxiv*, (2019).
103. M. Alexander *et al.*, Exosome-delivered microRNAs modulate the inflammatory response to endotoxin. *Nat Commun* **6**, 7321 (2015).
104. W. Li *et al.*, Rab27A regulates exosome secretion from lung adenocarcinoma cells A549: involvement of EPI64. *APMIS* **122**, 1080-1087 (2014).
105. Y. Zheng, E. C. Campbell, J. Lucocq, A. Riches, S. J. Powis, Monitoring the Rab27 associated exosome pathway using nanoparticle tracking analysis. *Exp Cell Res* **319**, 1706-1713 (2013).
106. A. Bobrie *et al.*, Rab27a supports exosome-dependent and -independent mechanisms that modify the tumor microenvironment and can promote tumor progression. *Cancer Res* **72**, 4920-4930 (2012).
107. S. Mohr *et al.*, Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* **19**, 958-970 (2013).
108. Y. Qin *et al.*, High-throughput sequencing of human plasma RNA by using thermostable group II intron reverse transcriptases. *RNA* **22**, 111-128 (2016).
109. X. Wu *et al.*, De novo sequencing of circulating miRNAs identifies novel markers predicting clinical outcome of locally advanced breast cancer. *J Transl Med* **10**, 42 (2012).

110. Z. Wei, A. O. Batagov, D. R. Carter, A. M. Krichevsky, Fetal Bovine Serum RNA Interferes with the Cell Culture derived Extracellular RNA. *Sci Rep* **6**, 31175 (2016).
111. O. Flores, E. M. Kennedy, R. L. Skalsky, B. R. Cullen, Differential RISC association of endogenous human microRNAs predicts their inhibitory potential. *Nucleic Acids Res* **42**, 4629-4639 (2014).
112. D. J. Gibbings, C. Ciaudo, M. Erhardt, O. Voinnet, Multivesicular bodies associate with components of miRNA effector complexes and modulate miRNA activity. *Nat Cell Biol* **11**, 1143-1149 (2009).
113. V. D'Souza, M. F. Summers, How retroviruses select their genomes. *Nat Rev Microbiol* **3**, 643-655 (2005).
114. C. M. Swanson, M. H. Malim, Retrovirus RNA trafficking: from chromatin to invasive genomes. *Traffic* **7**, 1440-1450 (2006).
115. L. Kleiman, tRNA(Lys3): the primer tRNA for reverse transcription in HIV-1. *IUBMB Life* **53**, 107-114 (2002).
116. D. J. Cha *et al.*, KRAS-dependent sorting of miRNA to exosomes. *Elife* **4**, e07197 (2015).
117. C. Coulouarn, V. M. Factor, J. B. Andersen, M. E. Durkin, S. S. Thorgeirsson, Loss of miR-122 expression in liver cancer correlates with suppression of the hepatic phenotype and gain of metastatic properties. *Oncogene* **28**, 3526-3536 (2009).
118. B. Wang, H. Wang, Z. Yang, MiR-122 inhibits cell proliferation and tumorigenesis of breast cancer by targeting IGF1R. *PLoS One* **7**, e47053 (2012).
119. J. Zhang *et al.*, Exosome and exosomal microRNA: trafficking, sorting, and function. *Genomics Proteomics Bioinformatics* **13**, 17-24 (2015).
120. M. Wang *et al.*, Emerging Function and Clinical Values of Exosomal MicroRNAs in Cancer. *Mol Ther Nucleic Acids* **16**, 791-804 (2019).
121. L. A. Gilbert *et al.*, CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451 (2013).
122. M. A. Horlbeck *et al.*, Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife* **5**, (2016).
123. L. A. Gilbert *et al.*, Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647-661 (2014).
124. C. C. Thoreen *et al.*, A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109-113 (2012).
125. D. C. Rio, Electrophoretic mobility shift assays for RNA-protein complexes. *Cold Spring Harb Protoc* **2014**, 435-440 (2014).
126. T. Peritz *et al.*, Immunoprecipitation of mRNA-protein complexes. *Nat Protoc* **1**, 577-580 (2006).
127. T. L. Bailey, C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36 (1994).
128. K. Mukherjee *et al.*, Reversible HuR-microRNA binding controls extracellular export of miR-122 and augments stress response. *EMBO Rep* **17**, 1184-1203 (2016).

129. B. Liang *et al.*, Characterization and proteomic analysis of ovarian cancer-derived exosomes. *J Proteomics* **80**, 171-182 (2013).
130. M. Demory Beckler *et al.*, Proteomic analysis of exosomes from mutant KRAS colon cancer cells identifies intercellular transfer of mutant KRAS. *Mol Cell Proteomics* **12**, 343-355 (2013).
131. F. S. Skottvoll *et al.*, Comparison of ultracentrifugation and a commercial kit for isolation of exosomes derived from glioblastoma and breast cancer cell lines. *bioRxiv*, (2018).
132. J. E. Stefano, Purified lupus antigen La recognizes an oligouridylate stretch common to the 3' termini of RNA polymerase III transcripts. *Cell* **36**, 145-154 (1984).
133. J. Rinke, J. A. Steitz, Association of the lupus antigen La with a subset of U6 snRNA molecules. *Nucleic Acids Res* **13**, 2617-2629 (1985).
134. S. L. Wolin, T. Cedervall, The La protein. *Annu Rev Biochem* **71**, 375-403 (2002).
135. R. V. Intine, S. A. Tenenbaum, A. L. Sakulich, J. D. Keene, R. J. Maraia, Differential phosphorylation and subcellular localization of La RNPs associated with precursor tRNAs and translation-related mRNAs. *Mol Cell* **12**, 1301-1307 (2003).
136. B. Cardinali, C. Carissimi, P. Gravina, P. Pierandrei-Amaldi, La protein is associated with terminal oligopyrimidine mRNAs in actively translating polysomes. *J Biol Chem* **278**, 35145-35151 (2003).
137. C. Crosio, P. P. Boyl, F. Loreni, P. Pierandrei-Amaldi, F. Amaldi, La protein has a positive effect on the translation of TOP mRNAs in vivo. *Nucleic Acids Res* **28**, 2927-2934 (2000).
138. M. Petz, N. Them, H. Huber, H. Beug, W. Mikulits, La enhances IRES-mediated translation of laminin B1 during malignant epithelial to mesenchymal transition. *Nucleic Acids Res* **40**, 290-302 (2012).
139. M. Petz, N. C. Them, H. Huber, W. Mikulits, PDGF enhances IRES-mediated translation of Laminin B1 by cytoplasmic accumulation of La during epithelial to mesenchymal transition. *Nucleic Acids Res* **40**, 9738-9749 (2012).
140. S. R. Baglio *et al.*, Sensing of latent EBV infection through exosomal transfer of 5'pppRNA. *Proc Natl Acad Sci U S A* **113**, E587-596 (2016).
141. G. S. Tan *et al.*, Expanded RNA-binding activities of mammalian Argonaute 2. *Nucleic Acids Res* **37**, 7533-7545 (2009).
142. R. J. Maraia, La protein and the trafficking of nascent RNA polymerase iii transcripts. *J Cell Biol* **153**, F13-18 (2001).
143. G. Sommer, C. Rossa, A. C. Chi, B. W. Neville, T. Heise, Implication of RNA-binding protein La in proliferation, migration and invasion of lymph node-metastasized hypopharyngeal SCC cells. *PLoS One* **6**, e25402 (2011).
144. R. Trotta *et al.*, BCR/ABL activates mdm2 mRNA translation via the La antigen. *Cancer Cell* **3**, 145-160 (2003).
145. F. Al-Ejeh, J. M. Darby, M. P. Brown, The La autoantigen is a malignancy-associated cell death target that is induced by DNA-damaging drugs. *Clin Cancer Res* **13**, 5509s-5518s (2007).
146. G. Sommer *et al.*, The RNA-binding protein La contributes to cell proliferation and CCND1 expression. *Oncogene* **30**, 434-444 (2011).

147. E. A. van Niekerk *et al.*, Sumoylation in axons triggers retrograde transport of the RNA-binding protein La. *Proc Natl Acad Sci U S A* **104**, 12913-12918 (2007).
148. D. Koppers-Lalic *et al.*, Nontemplated nucleotide additions distinguish the small RNA composition in cells from exosomes. *Cell Rep* **8**, 1649-1658 (2014).
149. D. G. Meckes, Jr. *et al.*, Modulation of B-cell exosome proteins by gamma herpesvirus infection. *Proc Natl Acad Sci U S A* **110**, E2925-2933 (2013).
150. P. Feist, A. B. Hummon, Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int J Mol Sci* **16**, 3537-3563 (2015).
151. V. Hung *et al.*, Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nat Protoc* **11**, 456-475 (2016).
152. D. W. Sirkis, R. E. Aparicio, R. Schekman, Neurodegeneration-associated mutant TREM2 proteins abortively cycle between the ER and ER-Golgi intermediate compartment. *Mol Biol Cell* **28**, 2723-2733 (2017).
153. A. Gorur *et al.*, COPII-coated membranes function as transport carriers of intracellular procollagen I. *J Cell Biol* **216**, 1745-1759 (2017).
154. V. Hung *et al.*, Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. *Elife* **6**, (2017).
155. S. Han *et al.*, Proximity Biotinylation as a Method for Mapping Proteins Associated with mtDNA in Living Cells. *Cell Chem Biol* **24**, 404-414 (2017).
156. S. S. Lam *et al.*, Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods* **12**, 51-54 (2015).
157. K. Bersuker *et al.*, A Proximity Labeling Strategy Provides Insights into the Composition and Dynamics of Lipid Droplet Proteomes. *Dev Cell* **44**, 97-112 e117 (2018).
158. C. Y. Chung *et al.*, In Situ Peroxidase Labeling and Mass-Spectrometry Connects Alpha-Synuclein Directly to Endocytic Trafficking and mRNA Metabolism in Neurons. *Cell Syst* **4**, 242-250 e244 (2017).
159. B. T. Lobingier *et al.*, An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells. *Cell* **169**, 350-360 e312 (2017).
160. S. Markmiller *et al.*, Context-Dependent and Disease-Specific Diversity in Protein Interactions within Stress Granules. *Cell* **172**, 590-604 e513 (2018).
161. T. Del Olmo *et al.*, APEX2-mediated RAB proximity labeling identifies a role for RAB21 in clathrin-independent cargo sorting. *EMBO Rep* **20**, (2019).
162. E. T. Williams, X. Chen, D. J. Moore, VPS35, the Retromer Complex and Parkinson's Disease. *J Parkinsons Dis* **7**, 219-233 (2017).
163. H. Zhang *et al.*, The Retromer Complex and Sorting Nexins in Neurodegenerative Diseases. *Front Aging Neurosci* **10**, 79 (2018).
164. P. Sanchez-Martin, M. Komatsu, p62/SQSTM1 - steering the cell through health and disease. *J Cell Sci* **131**, (2018).
165. J. Mejlvang *et al.*, Starvation induces rapid degradation of selective autophagy receptors by endosomal microautophagy. *J Cell Biol* **217**, 3640-3655 (2018).
166. J. L. Welton *et al.*, Proteomics analysis of bladder cancer exosomes. *Mol Cell Proteomics* **9**, 1324-1338 (2010).

167. M. He *et al.*, Hepatocellular carcinoma-derived exosomes promote motility of immortalized hepatocyte through transfer of oncogenic proteins and RNAs. *Carcinogenesis* **36**, 1008-1018 (2015).